THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXXII December, 1941 Number 9

THE RETENTION OF ORALLY PRESENTED MATERIALS

EDWARD M. SPENCER

State Teachers College,[1] Moorhead, Minn.

The main objectives of this investigation were (1) to study the effects of a written recognition test response upon the retention of orally presented materials, and (2) to compare the rate of forgetting and the effect of a recognition test response upon retention for two methods of presenting learning materials, visual and oral.

PROBLEM

The use of recall as an aid to retention has been emphasized by investigators in the field of retention. However, in general, there is a lack of experimental evidence showing the effect of recall on retention in situations approaching actual classroom practice. Recently, Spitzer[2] studied the effect of administering intervening tests on the retention of facts acquired by pupils from reading materials comparable to materials studied in the classroom. His results demonstrate clearly the benefits on the retention of facts by the response to intervening tests. Since information in the classroom is presented by oral as well as written language, the retention from oral language merits consideration. Will interpolated responses to recognition tests improve the retention of information acquired from oral presentation? This investigation was planned to study the effect of interpolated test responses on the retention of facts which children acquire from orally presented materials as well as to compare the effect of such tests on the retention of information acquired from oral presentation with the

[1] A digest of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Education of the State University of Iowa. The assistance of Dr. Ernest Horn, under whose direction the study was conducted, is gratefully acknowledged.

[2] Spitzer, H. F.: "Studies in Retention." *Jl. of Educ. Psych.*, Vol. xxx, 1939, pp. 641–656.

retention of information acquired from visual presentation. The data obtained in this study provide the information of the effect of interpolated tests on retention of facts from oral presentation. The comparison between oral and visual presentation is made possible by comparing the results from this investigation with the results obtained from the investigation by Spitzer.[1] The same materials were used in both studies and the subjects were, in both instances, sixth-grade pupils.

SUBJECTS AND MATERIALS

The subjects for this experiment were twenty-two hundred twelve sixth-grade pupils comprising seventy-two classes from thirteen school systems. The seventy-two classes were assigned at random by means of a table of random numbers to nine experimental groups of eight classes each. These groups were designated as Groups I to IX, inclusive.

The experimental materials used in this study were those prepared by Spitzer[1] for use in his study. The materials consist of two articles: The first, a 604-word article, entitled "Peanuts," designated *A*, was used in securing the initial measure of the pupils' ability to learn; the second article, five hundred seventy-seven words in length, entitled "Bamboo," designated *B*, was used in securing the measure of retention. The retention of facts was measured by a twenty-five item five-response multiple-choice test on each article. The within class correlation between the two tests was .70. For additional description of these materials, the reader is referred to the study by Spitzer.[1]

The Test *A* scores over Article *A* were used as initial measures of the learning ability of the pupils. The Test *B* scores over Article *B* were used as measures of the pupils' ability to retain the orally presented materials. By statistical methods, the Test *B* scores were corrected for differences in initial learning ability between the experimental groups as measured by Test *A*. The methods of analysis of covariance[2] were the statistical methods used to test the significance of the differences in the mean scores on Test *B* of the experimental groups. Since there were eight classes in each experimental group, each of the nine groups was treated as a sample of eight cases (eight class means).

---

[1] Spitzer, H. F.: *Op. cit.*

[2] Lindquist, E. F.: *Statistical Analysis in Educational Research.* Houghton Mifflin Co., Boston, 1940, Chapter VI.

PROCEDURE

Articles *A* and *B* were memorized by the experimenter and presented orally to all classes at a rate of approximately two words per second. The immediate tests were administered by the experimenter and all delayed tests by the classroom teacher.

Articles *A* and *B* were presented to the nine experimental groups on the first day of the experiment and all groups took a test over Article *A* the same day that the article was presented. The subjects in Group I were given the test over Article *B* prior to the presentation of the article. The subjects in Groups I, II, III, VIII, and IX took a test over Article *B* immediately following presentation of Article *B*. The subjects in Groups IV–VII, inclusive, took their first test over Article *B*, one, seven, fourteen, and sixty-three days after the article was presented. These groups, however, took an immediate test over Article *B*, the items in which had no relation to the items subsequently used in the later Test *B*. The scores on this test were not used in the analyses, but the test was given in order to keep the pupils from expecting a later test. All groups, excepting VII, took one or more delayed tests over Article *B*. The pattern of the time of administering tests is shown in Table I. The results on the sixty-three-day test were used as the criterion of retention.

TABLE I.—THE TESTING PROCEDURE FOR TEST *B*

Groups	Pretest	Immediate	1 day	7 days	14 days	21 days	63 days
I	B_1	B_2	....	....	....	....	B_3
II	..	B_1*	B_2*	....	....	B_3*	
III	..	B_1*	....	B_2*	....	....	B_3*
IV	..	....	B_1*	....	B_2*	....	B_3
V	..	....	....	B_1*	....	B_2*	B_3
VI	..	....	....	....	B_1*	....	B_2
VII	..	....	....	....	....	....	B_1*
VIII	..	B_1	....	....	....	B_2	
IX	..	B_1	....	....	....	....	B_3

* Comparable to the testing procedure for Spitzer's investigation.

B_1—the first time Test *B* was given.

B_2—the second time Test *B* was given.

B_3—the third time Test *B* was given.

B_1 for Group I is listed under pretest to show that it was given prior to the presentation of Article *B*.

The materials were not referred to after the initial presentation nor did the pupils know that there were to be delayed tests. The teachers were instructed not to discuss the articles or the tests with the pupils. Approximately five minutes each were required to present the articles and the pupils were given ten minutes to take each test.

By comparing the results on Test B_1 of Groups II, III, VIII, and IX, which took the test immediately, with the results from Groups IV, V, VI, and VII, which took Test B_1 one, seven, fourteen, and sixty-three days after learning, some idea may be gained of the rate of forgetting of Article *B*. This comparison is shown in Table II.

If the score on the immediate test (12.90) in Table II is taken as the measure of the amount learned, twenty-one per cent (difference between 12.90 and 10.13) was forgotten in one day. This is a slower

TABLE II.—MEAN SCORES ON TEST B_1 WHEN TAKEN AT DIFFERENT LENGTH INTERVALS AFTER LEARNING

Length of interval	Immediate	1 day	7 days	14 days	63 days
Groups II–III, VIII–IX	12.90[1]				
IV	.....	10.13			
V	.....	.....	8.48		
VI	.....	.....	....	7.24	
VII	.....	.....	....	....	6.74

[1] Average of the mean scores for the four groups.

initial rate of forgetting than has been reported by previous investigators. For example, Spitzer's[1] results show that forty-four per cent of the facts learned were forgotten in one day.

THE EFFECT OF A TEST RESPONSE ON RETENTION AS MEASURED BY THE CRITERION TEST

Test *B* was administered to Groups I, III, IV, V, VI, VII, and IX sixty-three days after learning. This test was the criterion test of retention for these groups. The time of the intervening tests over Article *B* varied for each group. (See Table I.) The initial (Test *A* means) and final criterion mean scores (Test *B*) for the seven groups are shown in Table III.

Table IV presents the analysis of covariance for these data. This table shows that the ratio ($F = 15.95$) between the reduced "between

[1] Spitzer, H. F.: *Op. cit.*

TABLE III.—THE INITIAL AND FINAL CRITERION MEAN SCORES FOR THE SEVEN GROUPS THAT TOOK THE CRITERION TEST

Groups	Test $A(x)$ mean	Test $B(y)$ mean	Test B adjusted mean
I	15.58	9.21	8.69
III	13.96	10.19	10.49
IV	14.57	8.94	8.94
V	14.36	7.97	8.07
VI	13.60	6.68	7.16
VII	14.76	6.74	6.64
IX	15.09	8.45	8.18
GM	14.56		

TABLE IV.—THE ANALYSIS OF COVARIANCE OF INITIAL AND CRITERION MEAN SCORES OF THE SEVEN EXPERIMENTAL GROUPS THAT TOOK THE CRITERION TEST

Source of variation	df	Sums of squares and products			Adjusted values		
		Σx^2	Σxy	Σy^2	y^2	df	Variance
Groups	6	21.39	9.50	79.92	75.67	6	12.61
Within groups (error)	49	100.97	51.02	63.98	38.20	48	.796
Total	55	122.36	60.52	143.90	113.87	54	

Test of significance of differences in adjusted means $F = 12.61/.796 = 15.95$ (2.30) (3.20).[1]

Estimated standard error of an adjusted mean $= \sqrt{.796/8} = .315$.

Estimated standard error of the difference between adjusted means (.315) (1.414) = .445.

For 48 degrees of freedom, a difference of (.445) (2.575) = 1.15 is necessary to be significant at the one per cent level.

r_{xy} (within groups) = .64 (Class means).

$b = 51.02/100.97 = .505$.

[1] The values (2.30) (3.20) are the values of F necessary for significance at the five per cent and one per cent levels, respectively.

groups" variance and the adjusted "within groups" variance is larger than that required for significance at the one per cent level.

The F test shows that the differences between adjusted means cannot be attributed entirely to chance. To determine the signifi-

cance of any particular difference between two means, the "t" test was used.[1] The formula for the value of "t" for the difference between two group means is

$$t = \frac{M_1 - M_2}{\text{est'd. } \sigma_{M_1 - M_2}}$$

in which $M_1 - M_2$ is the difference between any two group means and est'd. $\sigma_{M_1 - M_2}$ is the estimated standard error of the difference.

From this formula for the value of "t," it is possible to compute the minimum difference ($M_1 - M_2$) necessary for significance at the one per cent level for any pair of the adjusted means in Table III. A value of 1.15 was obtained by this method for the minimum difference necessary between any two adjusted means for significance at this level.

From the data represented in Tables III and IV several comparisons may be made between the experimental procedures represented by Groups I, III, IV, V, VI, VII, and IX, keeping in mind that a difference greater than 1.15 between any two adjusted means is necessary for significance at the one per cent level.

The distribution of B tests for Group $III_{0,7,63}$ with an adjusted mean of 10.49 was superior to the distributions for the other six groups, judged by the amount of retention as measured by the sixty-three-day criterion test.[2] The differences in the distributions for Group $I_{p,0,63}$, $IV_{1,14,63}$, $V_{7,21,63}$, and $IX_{0,63}$ produced no significant differences. The adjusted means for these four groups were 8.69, 8.94, 8.07, and 8.18 respectively. The pretest given to Group I prior to the presentation of the material had no significant effect on retention as measured by the criterion test. Also, the distribution of tests for Group IV apparently did not aid retention any more than did the single immediate test for Group IX.

These comparisons indicate that when recall is stimulated by an objective test, two tests, one immediately and one later, will be more effective in aiding retention than a single test (compare results on criterion test of Group III with IX), but unless the two tests come soon after learning, then a single immediate test is just as effective. (Compare results of Group IX with Groups IV and V.)

The results from Groups $I_{p,0,63}$, $III_{0,7,63}$, and $IV_{1,14,63}$ are in each case significantly superior to the results from either Groups $VI_{14,63}$ or VII_{63}. It should be noted that the first test was given Groups I, III,

[1] Lindquist, E. F.: *Op. cit.*, pp. 51–60.

[2] Subscripts refer to length of intervals between presentation of Article B and the taking of Test B.

and IV early in the experiment while the first test was given Groups VI and VII later in the experiment. Groups $I_{p,0,63}$, $III_{0,7,63}$, $IV_{1,14,63}$, $V_{7,21,63}$, and $IX_{0,63}$ are significantly superior to Group VII_{63}. The beneficial effects of interpolated test responses are also shown by this comparison as Group VII took no intervening tests prior to taking the criterion test.

RETENTION AFTER TWENTY-ONE DAYS

Groups $II_{0,1,21}$, $V_{7,21,63}$, and $VIII_{0,21}$, took Test *B* twenty-one days after learning. (See Table I.) For Groups II and VIII this was the

TABLE V.—THE INITIAL AND FINAL MEAN SCORES FOR THE THREE GROUPS THAT TOOK TEST *B* AFTER TWENTY-ONE DAYS

Groups	Test $A(x)$ mean	Test $B(y)$ mean	Test *B* adjusted mean
II	14.68	12.13	12.00
V	14.36	8.04	8.17
VIII	14.52	10.07	10.07
GM	14.52		

TABLE VI.—THE ANALYSIS OF COVARIANCE OF THE INITIAL AND FINAL MEAN SCORES OF THE THREE GROUPS THAT TOOK TEST *B* AFTER TWENTY-ONE DAYS

Source of variation	df	Sums of squares and products			Adjusted values		
		Σx^2	Σxy	Σy^2	y^2	df	Variance
Groups	2	.42	5.32	67.16	58.62	2	29.31
Within groups (error)	21	47.96	37.20	45.87	17.02	20	.85
Total	23	48.39	42.52	113.03	75.67	22	

Test of significance of differences in adjusted means. $F = 29.31/.85 = 34.$ (3.49) (5.45) are values of *F* necessary for significance at the five and one per cent levels.

Estimated standard error of the adjusted *B* means = .32.

Estimated standard error of a difference between adjusted means = .45.

Difference necessary between adjusted means for significance at the one per cent level = 1.35.

r_{xy} (within groups) = .79 (Class Means).

$b = 37.20/47.96 = .775$.

XUM

final test of retention. The initial and final means of these groups are shown in Table V. The analysis of covariance for these data is presented in Table VI.

A difference greater than 1.35 between adjusted means is necessary to be significant at the one per cent level. The differences between these adjusted means (see Table V) are all larger than 1.35. From these results, the most effective distribution of tests based upon the amount retained at the end of twenty-one days is the distribution for Group $II_{0,1,21}$, with Group $VIII_{0,21}$, as next, and Group $V_{7,21,63}$ as last. The extra test given to Group II (Mean 12.00) one day after learning produced significant results on later retention when compared with the results of Group $VIII_{0,21}$ (Mean 10.07) which had only the single immediate *B* test. The early tests taken by both groups produced superior retention when compared with the results of Group $V_{7,21,63}$. Here again as shown in Table II the importance of early tests is emphasized.

COMPARISON OF RETENTION AIDED BY ONE PREVIOUS TEST RESPONSE WITH RETENTION UNAIDED BY A TEST RESPONSE

Five comparisons may be made between retention when aided by one test response with retention when unaided by a test response. These comparisons are summarized in Table VII. In this table, B_2 indicates the results on Test *B* the second time whereas B_1 indicates the first time. The results on Test B_2 were aided by the previous response to Test B_1. The subscripts following the group number in column one indicate the intervals in days between learning and testing.

That a test response aids retention is shown by the fact that all differences shown in Table VII except one are significant at the one per cent level. This one instance is the comparison between Groups $VI_{14,63}$ and VII_{63}. The importance of a written test soon after learning is emphasized by this comparison. The first test for Group $VI_{14,63}$ came too late to offset the rapid initial rate of forgetting. Therefore, when retention was measured at the end of sixty-three days, the difference between the results for Groups $VI_{14,63}$ and VII_{63} was not significant.

The evidence from this investigation showing the importance of a test response as an aid to retention substantially corroborates the results obtained by Spitzer[1] in his investigation. Spitzer made seven

[1] Spitzer, H. F.: *Op. cit.*

comparisons between retention aided by one test response and retention unaided by a test response. In six of the seven comparisons, the differences were statistically significant. The one comparison in which the difference was not significant involved a situation comparable to the similar situation in this investigation in which the first test was not given soon enough after the initial learning period to offset the rapid initial rate of forgetting.

TABLE VII.—COMPARISON OF TEST *B* RESULTS FOR GROUPS DIFFERING BY ONLY ONE TEST RESPONSE

Group	Test	Mean	Difference
$II_{0.1.63}$	B_2	13.37[1]	
$IV_{1.14.63}$	B_1	10.13[1]	3.24
$III_{0.7.63}$	B_2	11.02[1]	
$V_{7.21.63}$	B_1	8.48[1]	2.54
$IV_{1.14.63}$	B_2	9.60[1]	
$VI_{14.63}$	B_1	7.24[1]	2.36
$IX_{0.63}$	B_2	8.18	
VII_{63}	B_1	6.64	1.54
$VI_{14.63}$	B_2	7.16	
VII_{63}	B_1	6.64	.52

[1] Unadjusted statistically to allow for differences in learning ability between the groups.

COMPARISON OF VISUAL AND ORAL PRESENTATION

In this section reference to oral presentation will refer to this investigation and visual presentation to Spitzer's[1] investigation. Table I shows the testing procedures that were comparable for both investigations.

To facilitate the comparison of the two methods of presentation, certain pupil scores on Test *A* in this investigation were eliminated from each group (I–IX) so that the means and standard deviations of the remaining scores were approximately the same for each group. In this manner the nine groups were equated on the basis of the scores made on Test *A*. This procedure was used by Spitzer to equate the groups in his study. Since all groups were equated on the basis of

[1] *Ibid.*

the results on Test *A* and since Articles *A* and *B* are similar, it is assumed that all groups may be expected to learn approximately equal amounts from a presentation and that the groups possess equal ability to retain the effects of the presentation. According to this assumption, direct comparisons of results from group to group may be made. After certain scores were eliminated, the means and standard deviations of the remaining scores were computed for each group. These results are shown in Table VIII.

TABLE VIII.—MEANS AND STANDARD DEVIATIONS OF SCORES ON TEST *A* OF EQUATED GROUPS

Groups	*N*	Mean	SD
I	224	14.59	4.34
II	211	14.52	4.26
III	206	14.55	4.28
IV	270	14.57	4.34
V	237	14.56	4.32
VI	196	14.52	3.93
VII	230	14.57	4.20
VIII	202	14.58	3.98
IX	210	14.55	4.16
Total	1984	14.56	4.32

The range of means of these equated groups is 14.52 to 14.59. In Spitzer's study the range was 15.00 to 15.05.

In order to compare the rate of forgetting and the effect of intervening test responses on retention for the two modes of presentation, a comparison of the results on Test *B* for comparable time intervals between learning and testing is shown in Table IX. The means for oral presentation in this table are the means on Test *B* of the equated groups referred to in Table VIII.

The data in Table IX are shown graphically in Fig. 1. The two main lines represent the curves of retention for oral and visual presentation. The points on these lines represent the mean performance of the different groups on their first attempt on Test *B*. For example, the points designated B_1 represent the mean performance on Test *B* the first time the test was taken one, seven, fourteen, and sixty-three days after the learning period. Since the different groups were equalized according to their performance on Test *A*, these curves assume that the facts learned by the groups which took B_1 immedi-

XUM

TABLE IX.—TEST *B* RESULTS OF COMPARABLE TIME INTERVALS BETWEEN LEARNING AND RECALL FOR VISUAL AND ORAL PRESENTATION

Test	Oral	Visual	*N* oral	*N* visual
B_1-0	13.27	13.23	211	286
B_2-1	13.27	13.07	207	284
B_3-21	12.11	12.18	201	266
B_1-0	12.75	13.20	216	338
B_2-7	11.69	11.84	204	335
B_3-63	10.72	10.74	193	312
B_1-1	10.18	9.56	265	367
B_2-14	9.58	8.93	259	349
B_1-7	8.55	7.87	226	337
B_3-21	8.11	8.15	213	323
B_1-14	7.86	6.97	194	371
B_1-63	6.63	6.38	215	350

B_1-0 means the first *B* test administered immediately after learning.
B_2-1, the second test given one day later, etc.

ately would have been forgotten at the rate indicated by the main lines if they had not been given an intervening test. The secondary

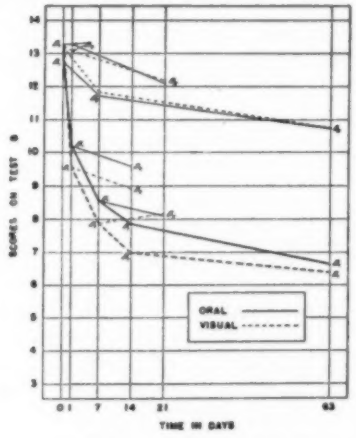

FIG. 1 CURVES OF RETENTION WHEN THE AMOUNT RETAINED IS EXPRESSED IN RAW SCORES

lines show the beneficial effect on retention of a response to a recognition test. Thus the vertical distance between a point on the main

lines and a point on the secondary lines represents the effect of a test response on retention. This effect on retention for the two methods of presentation may be noted by comparing the different points on the main lines with similar points on the secondary lines. The two main curves emphasize the rapid initial rate of forgetting and the secondary lines the beneficial effect of intervening tests on retention, as well as the importance of these tests coming soon after the learning period. Although these curves show that the rate of forgetting is somewhat slower for oral presentation, the general trend is the same for both methods of presentation.

The comparison between the two investigations of the effects of intervening test responses on retention has been made by showing the differences between mean scores on Test *B* for comparable testing situations. (See Table IX.) In view of the fact that these studies were conducted during different years with different subjects, no attempt has been made to evaluate the significance of these differences.

EFFECT OF ITEM DIFFICULTY ON RETENTION

From an item analysis of Test *B* papers, the data on forgetting of individual items for both presentations shown in Table X were obtained. The first column represents the per cent of correct answers on the test administered to obtain a measure of previous knowledge.

The last five columns show the per cent of correct responses on each item at various intervals of time after learning.

A study of Table X will show a rather high degree of similarity between the rates of forgetting on each item for both presentations. A comparison of the curves of retention for a few of the items is shown in Figs. 5, 6, 7, and 8. The points on the curves represent the per cent of correct responses on one, seven, fourteen, and sixty-three days after learning. The curves for items 4, 11, 14, 16, 17, and 19 show a comparison of curves for items of approximately the same difficulty for both presentations. The two curves for the same item are rather similar. The other curves presented also emphasize the similarity of rates of forgetting.

Limitations.—The findings of this study are subject to limitations, the most important of which are as follows: (1) The type of test (recognition) used. (2) The tests were repeated. (3) The subjects had no opportunity of referring to the material after the initial learning period. (4) The pupils did not know whether their responses were correct or

incorrect. Neither the corrected tests nor the reading materials were made available to the students, so they had no opportunity to relearn items shown to be wrong on any previous test. What would have been the influence on retention had the tests been corrected, placed in the hands of the pupils, and used as a guide for review is hard to say.

TABLE X.—PER CENT OF CORRECT RESPONSES FOR EACH OF THE ITEMS ON TEST *B* FOR ORAL AND VISUAL PRESENTATION

Item No.	Previous knowledge		Immediate		1 day		7 days		14 days		63 days	
	O	*V*	*O*	*V*	*O*	*V*	*O*	*V*	*O*	*V*	*O*	*V*
1	5	10	36	40	21	28	20	21	19	19	22	14
2	4	14	65	61	42	30	25	20	24	15	23	15
3	37	60	86	81	59	62	54	52	43	45	38	50
4	16	18	80	80	52	56	37	38	35	36	27	28
5	37	56	53	55	43	42	49	40	43	37	45	46
6	47	66	70	76	60	57	69	57	65	55	64	61
7	15	17	47	43	40	40	26	33	31	29	34	25
8	6	9	36	32	31	24	27	18	31	19	30	19
9	18	26	42	75	35	60	39	53	45	46	39	40
10	8	11	52	55	29	24	23	21	21	20	18	13
11	22	38	72	72	40	50	22	29	21	23	29	31
12	3	6	37	29	27	22	27	19	22	16	13	10
13	14	19	62	67	35	30	22	24	29	19	24	26
14	10	20	74	77	70	64	64	57	62	44	50	33
15	4	9	68	73	52	46	25	34	24	24	21	14
16	3	7	42	41	31	28	30	27	24	17	12	14
17	3	8	41	40	28	28	26	22	23	18	16	13
18	3	4	29	37	14	20	16	19	14	17	9	9
19	3	4	66	68	32	39	31	32	22	22	13	9
20	10	21	52	53	44	39	41	32	33	34	33	32
21	4	14	23	63	10	27	16	16	11	14	12	13
22	4	9	23	24	20	17	16	14	16	19	6	17
23	22	34	68	73	59	56	64	54	56	52	49	50
24	11	22	27	35	26	31	35	27	27	27	24	28
25	7	10	32	31	21	21	24	21	25	19	10	20

However, in the light of other investigations upon the influence of knowledge of results on learning, it is safe to assume that had the pupils possessed a knowledge of results on previous tests, their scores would have been higher on subsequent tests, thereby resulting in increasing the beneficial results of tests on retention.

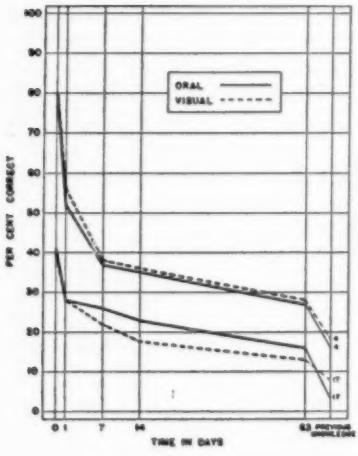

FIG. 5 CURVES OF RETENTION FOR ITEMS 4 & 17 (PER CENTS UNCORRECTED)

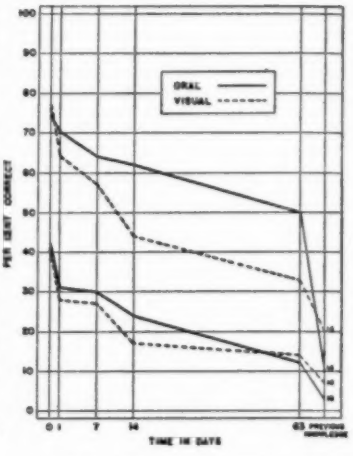

FIG. 6 CURVES OF RETENTION FOR ITEMS 14 & 16 (PER CENTS UNCORRECTED)

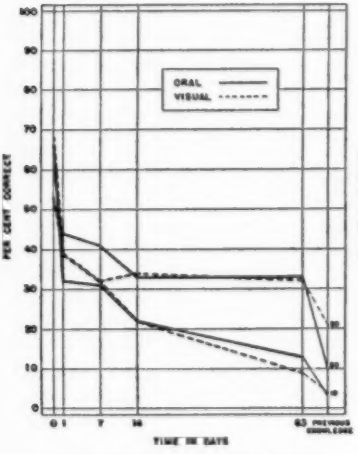

FIG. 7 CURVES OF RETENTION FOR ITEMS 19 & 20 (PER CENTS UNCORRECTED)

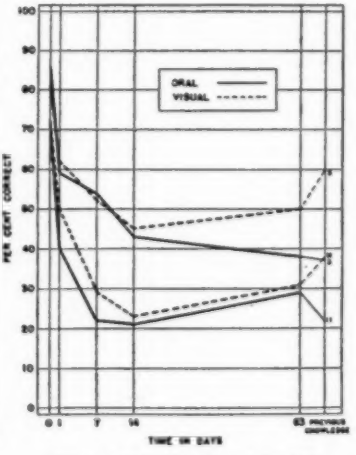

FIG. 8 CURVES OF RETENTION FOR ITEMS 3 & 11 (PER CENTS UNCORRECTED)

CONCLUSIONS AND IMPLICATIONS

(1) Recall stimulated by a written recognition test is beneficial to the retention of information presented orally and is most effective in retarding forgetting when the test is given immediately after learning.

(2) A response to two intervening tests soon after learning results in significantly better retention than does a response to one intervening test.

(3) The effectiveness of a written recognition test in retarding forgetting is not limited to one method of presentation but it may profitably be used to aid the fixation of ideas learned from oral instruction as well as from reading.

BIBLIOGRAPHY

1. Carver, M. E.: "Listening *vs.* Reading." Reported in Cantril & Allport. *The Psychology of Radio*, Chapter IX, Harper & Brothers, New York, 1935.
2. Corey, Stephen M.: "Learning from Lecture *vs.* Learning from Readings." *Jl. of Educ. Psych.*, Vol. XXV, 1934, pp. 459–470.
3. DeWick, H. N.: "The Relative Recall Effectiveness of Visual and Auditory Presentation of Advertising Material." *Jl. of Applied Psych.*, Vol. XIX, 1935, pp. 245–264.
4. Elliot, F. R.: "Memory for Visual, Auditory, and Visual-Auditory Material." *Archives of Psych.*, No. 199, 1936, pp. 58.
5. Erickson, C. I., and King, Irving: "A Comparison of Visual and Oral Presentation of Lessons in the Case of Pupils from the Third to the Ninth Grades." *School and Society*, Vol. VI, No. 136, 1917, pp. 146–148.
6. Greene, E. B.: "The Relative Effectiveness of Lecture and Individual Reading as Methods of College Teaching." *Genetic Psych.*, 1928, Mono. IV, pp. 457–563.
7. Lindquist, E. F.: *Statistical Analysis in Educational Research.* Houghton, Mifflin, Co., Boston, 1940.
8. Russell, R. D.: *The Relative Effectiveness of Presenting Verbal Material Visually and Orally as Measured by the Amount of Recall.* Doctor's dissertation, State University of Iowa, 1923.
9. Spitzer, H. F.: "Studies in Retention." *Jl. of Educ. Psych.*, Vol. XXX, 1939, pp. 641–656.
10. Worcester, D. A.: "Memory by Visual and by Auditory Presentation." *Jl. of Educ. Psych.*, Vol. XVI, 1925, pp. 18–27.
11. Young, W. E.: *The Relation of Reading Comprehension and Retention to Hearing Comprehension and Retention.* Doctor's dissertation, State University of Iowa, 1930.

XUM

STUDENT EVALUATION OF KNOWING COLLEGE APTITUDE TEST SCORE

R. K. COMPTON

Head Department of Psychology, South Dakota State College, Brookings, S. Dak.

The purpose of this study is to determine what evaluation students put upon the procedure followed by a number of institutions of informing the student of his percentile score on a test of general mental ability. What effect does the knowledge have on the student's scholarship aim, scholarship achievement, emotionality, course of study, and vocational choice? Are the effects related to the factors; of (*a*) relationship of score to student's self-estimated ability, (*b*) actual tested ability, and (*c*) amount of counseling given on basis of test score?

The subjects were one thousand three hundred thirty-one students from twenty-seven colleges who had been informed of their percentile scores on the American Council on Education Psychological Examination, 1939 edition. Most of the students replying were freshmen who had received their scores soon after taking the test; the remainder were first term sophomores. The student had had the knowledge of his score for several months preceding the reporting of his evaluations.

The body of the opinionnaire used to secure reactions of the students was as follows:

1. My score was { higher than I expected. / lower than I expected. / about what I expected. }

2. The knowledge of my mental ability score has { raised / lowered / not affected } my scholarship aim.

3. My scholarship achievement this year will likely be { higher / lower / unchanged } as a result of knowing my mental ability.

4. My attitude toward self (self-regard) has { improved / lowered / not changed } because of the knowledge of my score.

5. The knowledge of my mental score has { improved / lowered / not affected } my emotionality. (Moods, attitude toward work, morale.)

6. In the light of my mental ability score I am { thinking of changing / continuing with } my course of study.

7. In the light of my mental ability score I am { thinking of changing / continuing with } my vocational choice.

8. In general, I judge the knowledge of my tested ability to be { beneficial. / detrimental. / of little significance. / of uncertain value.

The opinionnaire was administered by college personnel officers in the respective twenty-seven institutions coöperating in the study.

These officers also provided two items of information concerning each student: (1) His national percentile score on the test, and (2) the amount of counseling the student received. This latter information was presented by using the code:

C_0..... Not counseled, except for a statement of the meaning of "percentile score."
C_1..... Counseled in group situation only.
C_2..... Individual conference, but rather a minimum of guidance given.
C_3..... Individual conference wherein the student's tested ability was related to such factors as course of study, vocational choice, personality adjustment, etc. (Use C_3 if you feel the counseling was fairly adequate.)

The data of this study are summarized (1) for the group as a whole, (2) in respect to the student's self-estimate of ability as provided in item one of the opinionnaire, (3) in relation to level of ability, and (4) in relation to the amount of counseling.

RESULTS FOR THE GROUP AS A WHOLE

The similarity of the findings in respect to the four attributes—scholarship aim, scholarship achievement, self-regard, and emotionality—permits them to be summarized together. It will be observed from Table I that approximately one student in three reports that the knowledge of his mental ability score resulted in improvement of scholarship aim and that this same approximate ratio holds for scholarship achievement, self-regard and emotionality. For each of these four attributes about three students in every five report no effect resulting from being informed of intelligence test scores. Negative effects are reported by some students. From four to six per cent experience a detrimental influence on the two scholarship traits and on emotionality. In the case of self-regard this negative effect increases to eleven and seven-tenths per cent.

Does the knowledge of tested ability have any relation to the planned course of study and to the vocational choice of the student? Table I summarizes the reports showing that eight and seven-tenths per cent of students contemplate a different course of study, and that six and five-tenths per cent are making a new vocational choice.

The evaluation placed upon knowing tested ability is greater when the effects are considered as a whole than when they are related to specific attributes. Whereas a student may have been negatively influenced in one or two aspects, more than half of the students (fifty-four and five-tenths per cent) state that in general they judge the knowledge of ability to be beneficial. Only two and one-tenth per cent conclude that the general result is detrimental. The remainder

TABLE I.—NUMBER AND PERCENTAGE OF STUDENTS MAKING EACH RESPONSE TO THE OPINIONNAIRE[1]

Item	Number responding	Variable	Number	Percentage
1	1310	Self-estimate		
		Score above	292	.223
		Score below	410	.313
		Score equal	608	.464
2	1307	Scholarship Aim		
		Raised	453	.347
		Lowered	67	.051
		Not affected	787	.602
3	1311	Scholarship Achievement		
		Higher	401	.306
		Lower	56	.043
		Unchanged	854	.651
4	1305	Self-regard		
		Improved	375	.287
		Lowered	153	.117
		Not changed	777	.595
5	1323	Emotionality		
		Improved	457	.345
		Lowered	81	.061
		Not affected	785	.593
6	1211	Course of Study		
		Thinking of changing	105	.087
		Continuing with	1106	.913
7	1303	Vocational Choice		
		Thinking of changing	85	.065
		Continuing with	1218	.934
8	1293	General Evaluation		
		Beneficial	705	.545
		Detrimental	27	.021
		Of little significance	271	.210
		Of uncertain value	290	.224

[1] The probable error of all twenty-three percentages is less than .01. All proportions are statistically significant.

of the students reporting, found the procedure was "of little significance" or "of uncertain value."

RESULTS FOR STUDENTS WHOSE SCORES WERE HIGHER OR LOWER THAN THEY HAD ANTICIPATED

Preliminary study gave evidence that a student's reactions to being informed of his test score depended in part on the relationship of his actual score to his self-estimated intelligence. Consequently, in this study the student was asked in Item 1 to indicate whether his score was higher, lower, or equal to what he judged his ability to be.

Table II summarizes the responses of two hundred ninety-two students whose scores were higher than they had anticipated, four hundred ten whose scores were lower, and six hundred eight whose

TABLE II.—RESULTS IN PERCENTAGES FOR STUDENTS GROUPED ACCORDING TO WHETHER SCORE WAS HIGHER[1] THAN, LOWER THAN, OR EQUAL TO SELF-ESTIMATED ABILITY

Attribute	Scores higher than expected $N = 292$	Scores lower than expected $N = 410$	Approximately the same $N = 608$
1. Scholarship aim			
(*a*) Raised	.543	.309	.273
(*b*) Lowered[2]	.014	.119	.280
2. Scholarship achievement			
(*a*) Likely to be higher	.404	.312	.356
(*b*) Likely to be lower	.028	.062	.037
3. Self-regard			
(*a*) Improved	.443	.192	.284
(*b*) Lowered	.024	.294	.059
4. Emotionality			
(*a*) Improved	.452	.335	.308
(*b*) Lowered	.027	.114	.043
5. Course of study			
(*a*) Plan to change	.090	.102	.074
6. Vocational choice			
(*a*) Plan to change	.053	.074	.065
7. General evaluation			
(*a*) Beneficial	.707	.439	.552
(*b*) Detrimental	.007	.052	.007

[1] The obtained proportions are all more than four times their respective probable errors with the exceptions of the measures .014, .007 and .007.

[2] The remaining percentages of students in each category responded that the knowledge of test scores had no effect. The same is true for each of the categories in Tables III and IV.

XUM

scores were equal to the student's self-appraisal. Rather marked differences were found for the three groups. Those who had underestimated their intelligence found the knowledge of their test score beneficial more frequently than did those who had overestimated their ability. Likewise, the detrimental influences of scores were quite pronounced when the scores were lower than the student anticipated. These two observations hold in respect to scholarship aim and achievement, emotionality, and most of all in respect to self-regard. Comparisons of these two groups—the underestimators and the overestimators—may be made from the data presented in Table II. For example fifty-four and three-tenths per cent of those whose scores were higher than they anticipated (the underestimators) reported that they raised their aim. Only one and four-tenths per cent of the underestimators lowered their aim, while twelve per cent of the overestimators had a lesser aim after receiving their score. The difference in response of these two groups is greatest in respect to self-regard. Of the overestimators twenty-nine and four-tenths per cent experienced less regard for themselves. Depreciation in self-regard for the underestimators occurred in only two and four-tenths per cent of the cases. The general evaluation by the underestimators was seventy and seven-tenths per cent beneficial and seven-tenths per cent detrimental as compared with the overestimators forty-three and nine-tenths per cent beneficial and five and two-tenths per cent detrimental. Comparison of the two groups may be made in like manner for all attributes reported.

The extent to which a student's reaction to being informed of his test score varies with the relationship of his subjectively estimated ability to his tested ability, may be expressed by the coefficient of contingency.[1]

Subjective-objective factor and:

Scholarship aim	.28
Scholarship achievement	.14
Self-regard	.35
Emotionality	.18
Course of study	.04
Vocational choice	.28
General effects	.23

RESULTS FOR STUDENTS OF FIVE ABILITY GROUPS

The extent to which the student's responses are related to his level of ability is indicated in Table III. The percentages of students

[1] By the Chi-square Test all seven of the coefficients are significant beyond .01.

TABLE III.—RESULTS IN PERCENTAGES FOR STUDENTS OF FIVE ABILITY GROUPS[1]

Attribute	81–100 $N = 311$	61–80 $N = 220$	41–60 $N = 226$	21–40 $N = 236$	0–20 $N = 191$
1. Self-estimate					
(*a*) Score higher than expected	.471	.215	.120	.105	.037
(*b*) Score lower than expected	.084	.233	.396	.459	.560
2. Scholarship aim					
(*a*) Raised	.431	.308	.379	.269	.304
(*b*) Lowered	.016	.014	.048	.075	.131
3. Scholarship achievement					
(*a*) Likely to be higher	.326	.264	.298	.275	.359
(*b*) Likely to be lower	.016	.009	.040	.081	.088
4. Self-regard					
(*a*) Improved	.348	.255	.241	.261	.333
(*b*) Lowered	.045	.086	.162	.179	.177
5. Emotionality					
(*a*) Improved	.373	.317	.309	.347	.417
(*b*) Lowered	.041	.036	.071	.084	.073
6. Course of study					
(*a*) Plan to change	.056	.040	.086	.120	.192
7. Vocational choice					
(*a*) Plan to change	.032	.041	.085	.064	.118
8. General evaluation					
(*a*) Beneficial	.611	.577	.553	.504	.523
(*b*) Detrimental	.010	.014	.022	.017	.046

[1] All but six of the above proportions are statistically significant.

reporting improvement in scholarship aim, scholarship achievement, self-regard and emotionality are not, as a whole, widely variant for the five ability groups. In general from one-fourth to one-third of the students within each of the five quintiles report improvement in these four attributes. The range of the twenty proportions (constituted by five ability groups in respect to the four attributes) is twenty-four and one-tenth per cent to forty-three and one-tenth per cent. While there are a number of statistically significant differences, most of them are relatively small and not directly proportional to the ability factor. The percentages of students reporting a lowering of scholarship aim, scholarship achievement, self-regard and emotionality are much more variant than for the groups reporting benefit. The range of the twenty proportions is nine-tenths per cent to seventeen and seven-tenths per cent. The detrimental effects are reported much more frequently in the low ability groups, and especially for the attribute of self-regard. A considerable proportion of the low ability group anticipate

XUM

a change in course of study (nineteen and two-tenths per cent) and likewise a change of vocational choice (eleven and eight-tenths per cent). The proportion of students in the other four ability groups influenced in these two attributes is much smaller and not, as a whole, directly proportional to the ability factor. Similarly the proportion of students in each of the five ability groups reporting general benefit from knowing score, and the proportions reporting detrimental effects seem not to be significantly related to the ability factor.

The contingency coefficients between level of ability and each of the eight attributes of the opinionnaire express the degree of relationship. The coefficients are as follows:[1]

Ability and	
Self-estimate	.43
Scholarship aim	.23
Scholarship achievement	.17
Self-regard	.18
Emotionality	.30
Course of study	.16
Vocational choice	.12
General evaluation	.11

RESULTS FOR FOUR GROUPS OF STUDENTS DIVIDED ACCORDING TO AMOUNT OF COUNSELING GIVEN

The extent to which the student's responses are related to the amount of counseling he is given in connection with his mental ability score is presented in Table IV. There are several significant differences of proportions which suggest that a little counseling is a dangerous thing. Larger proportions of those students given no counseling than those given group counseling report benefit in each of the four attributes—scholarship aim, scholarship achievement, self-regard and emotionality. Also, as an example of the evidence for this principle, fifteen and three-tenths per cent of the students guided only in the group situation report a lowered self-regard, while only six and eight-tenths per cent of those given no counseling are negatively affected. On the other hand the negative results for scholarship aim, scholarship achievement and emotionality are not as great when group counseling is given as when none at all is given. The differences

[1] By the Chi-square Test the first seven coefficients are significant beyond .01. There are twenty chances in one hundred that the obtained correlation (.11) between ability and general evaluation could have arisen from fluctuations of random sampling.

between the two procedures of group counseling and giving a minimum of guidance in individual conferences are small and in most of the proportions not statistically significant. The greatest number of students are benefited in the four attributes when the amount of counseling given in individual conference is considered by the counselor to be fairly adequate. The proportions of students negatively affected when the counseling is adequate are a little larger than the proportions when only a minimum of guidance is given, except in respect to emotionality, in which case the proportion is less. In general, then, the effect of counseling is to produce more definite reactions. Fewer students report no effects from knowing scores when counseling is adequate. In respect to the two attributes, course of study and vocational choice, the proportion of students contemplating a change is directly proportional to the amount of counseling given with marked differences occurring when the counseling is adequate. The general

TABLE IV.—RESULTS IN PERCENTAGES FOR STUDENTS GROUPED ACCORDING TO THE AMOUNT OF COUNSELING RECEIVED[1]

Counseling	C_0 $N = 74$	C_1 $N = 451$	C_2 $N = 299$	C_3 $N = 362$
1. Self-estimate				
(*a*) Score higher than expected	.370	.171	.217	.246
(*b*) Score lower than expected	.274	.391	.278	.282
2. Scholarship aim				
(*a*) Raised	.338	.311	.301	.429
(*b*) Lowered	.108	.049	.046	.052
3. Scholarship achievement				
(*a*) Likely to be higher	.297	.241	.288	.375
(*b*) Likely to be lower	.068	.034	.033	.087
4. Self-regard				
(*a*) Improved	.301	.226	.311	.362
(*b*) Lowered	.068	.153	.104	.113
5. Emotionality				
(*a*) Improved	.324	.308	.331	.411
(*b*) Lowered	.095	.064	.089	.041
6. Course of study				
(*a*) Plan to change	.041	.055	.068	.133
7. Vocational choice				
(*a*) Plan to change	.000	.043	.060	.106
8. General evaluation				
(*a*) Beneficial	.479	.472	.549	.468
(*b*) Detrimental	.014	.019	.040	.014

[1] All but six of the above proportions are statistically significant.

XUM

evaluation of guidance on the basis of test score is definitely greatest when counseling is adequate.

The coefficients of contingency between amount of counseling and each of the eight attributes of the opinionnaire are as follows:[1]

Amount of counseling and	
Self-estimate	.27
Scholarship aim	.14
Scholarship achievement	.17
Self-regard	.42
Emotionality	.11
Course of study	.13
Vocational choice	.13
General effects	.17

GENERAL CONCLUSIONS

Within the limits of the validity of self-judgments as obtained in this opinionnaire the following general conclusions are drawn:

(1) Approximately one student in three is benefited by the knowledge of his tested ability in the specific attributes—scholarship aim, scholarship achievement, self-regard and emotionality.

(2) Over fifty per cent of students consider that, taken as a whole, the knowledge of mental ability test score is beneficial.

(3) Detrimental effects on scholarship aim, scholarship achievement and emotionality are experienced in from four to six per cent of students, and in self-regard in eleven and seven-tenths per cent of the cases.

(4) A small but significant number of students change their course of study and their vocational choice when informed of their mental ability.

(5) Students who have estimated their ability to be higher than that which they test receive slightly less benefit than do those who have underestimated their intelligence.

(6) The relationship between level of ability and the benefit the student receives is small. Negative effects on students of low ability are not pronounced.

(7) Students value the procedure of receiving objective rating in intelligence more highly when they are adequately counseled.

(8) The data of this study suggest that college freshmen are sufficiently mature to profit by the knowledge of their mental ability test score.

[1] By the Chi-square Test all seven of the coefficients are significant beyond .01.

XUM

A STUDY OF THE EFFECT OF VARIOUS TEACHING PROCEDURES UPON OBSERVED GROUP ATTENTION IN THE SECONDARY SCHOOL[1]

R. W. EDMISTON

Miami University

AND

R. W. BRADDOCK

Waynesville, Ohio

TEACHING PROCEDURE AND ATTENTION

Group control of attention is a significant index of teaching efficiency.[2] The failure of a teacher to control the attention of pupils obviously will result in their failure to acquire abilities that have been specified as immediate objectives.[3]

In an unpublished master's thesis in the library of Miami University, Schaffer[4] pointed to the greater attention obtained by the demonstration than the laboratory procedure and to the agreement between charted observable attention and measured achievement in chemistry as indicative of correlation between observed attention and learning results. Bjarnason[5] in studying class size and attention concluded that groups' attention was affected more by the technique employed by the teacher than by the size of the class.

The above conclusions presented the questions: (1) Does observed group attention vary with different teaching procedures? (2) What implications does any variation offer for the improvement of teaching?

CHARTING ATTENTION UNDER VARIOUS TEACHING PROCEDURES

Twelve Ohio schools coöperated in this study. The method used in obtaining the attention scores was that advanced by Morrison.[6]

[1] A digest of Mr. Braddock's master's thesis which is in the library of Miami University, Oxford, Ohio.

[2] Knudson, C. W.: *Evaluation and Improvement of Teaching.* Doubleday, Doran and Company, Garden City, N. Y., 1932, p. 280.

[3] *Ibid.*, page 19.

[4] Schaffer, Mack: *Laboratory and Lecture-Demonstration Methods of Teaching High School Chemistry.* An unpublished thesis. Miami University Library, Oxford, Ohio, 1939.

[5] Bjarnason, L.: "Relation of Class Size to Control of Attention." *Elementary School Journal,* March, 1930, pp. 26, 36, 44.

[6] Morrison, H. C.: *The Practice of Teaching in the Secondary School.* The University of Chicago Press, Chicago, Ill., 1925, p. 119.

To measure group control of attention, Morrison suggested that the observer station himself near the front of the room so that he can look into the faces of the pupils yet be out of the natural line of their vision. The observer then records, minute by minute, the number in the class in attention. The physical attitude, the expression of the eyes, and pupil activity are the factors which Morrison considered as setting the attentive apart from the inattentive.

Blume,[1] Knudson,[2] and others have found the method reliable. In a number of studies carried out by Knudson, it was found that, with practice, the scores of most observers were in close agreement when each observer scored the same class. Blume likewise reported the method reliable when the observer was prepared to use it.

In comparison of this method with score cards and informal estimates as means for measuring attention, Shannon[3] discovered the attention score to be much more reliable than either of the other two methods.

The preliminary procedure consisted in an investigation of the teaching procedures in use in the twelve high schools and the training of observers from the volunteers who were principals and students. Unless agreements of more than ninety per cent were obtained with two other scorers in preliminary trials, the volunteer was not used. Pupils were found to create a less artificial situation in the opinion of the principals. After taking seats lending themselves to observation, the trained observers seldom disagreed, and most disagreements were in terms of a minute variation on the time of change in procedure. Disagreement or notations of irregularities and disturbances, like passing of fire department, resulted in the discard of the charted attention. The definitions of teaching procedures and directions for scoring derived from the preliminary procedure by and for the coöperating group were as follows.

DEFINITIONS OF TERMS

The term, teaching procedures, refers to the methods used by the teacher throughout the period of recitation to obtain the objectives of a particular

[1] Blume, C. E.: *Technique in the Measuring of Pupil Attention.* The Second Yearbook of the Department of Supervisors and Directors of Instruction, Chap. III, New York, Bureau of Publications, Teachers College, Columbia University, 1929.

[2] Knudson, C. W.: *Op. cit.*, pp. 323–332.

[3] Shannon, J. R.: "A Comparison of the Means for Measuring Efficiency in Teaching." *Journal of Educational Research*, Vol. XXIX, March, 1936, pp. 501–508.

lesson. The methods considered are the following: Laboratory, demonstration, lecture, reports, general discussion, teacher-dominated general discussion, pupil-dominated general discussion, rapid-fire question-answer, and workbook.

By the laboratory method is meant the use of a room equipped with scientific instruments to carry out experiments in general science, biology, botany, zoology, chemistry, or physics. It is obvious that some of this equipment must be used if the method is to be termed laboratory.

The demonstration method refers to an explanation or demonstration made by either the teacher or some other designated individual in which a principle or process is exemplified to the class by using apparatus, or any other means that lends itself to the object of the demonstration.

For the purpose of this study, the lecture method will be considered as the domination of the class period by the teacher for five minutes or more continuously for the purpose of presentation or explanation of material to the class. During the period of the lecture there would be no discussion on the part of the pupils.

General discussion refers to the procedure by which the teacher asks leading questions and thus leads the members of the class to give expression to their understanding of the topic or questions. This procedure might be either pupil- or teacher-dominated. Questions might go from pupil to pupil, or from pupil to teacher, as well as from teacher to pupil. This procedure is often characterized by criticisms, objections, and arguments among the members of the class and between the class and teacher. When the pupils do most (seventy-five per cent) of the discussing, the procedure is termed pupil-dominated; and when the discussion is dominated by the teacher, it is so termed. In those cases in which there is no clear domination by either teacher or students, the procedure is termed general discussion.

Procedure by reports refer to the giving of oral reports to the class by students.

The rapid-fire question-answer procedure is self-explanatory. It is understood that the teacher is to ask the questions under this procedure.

Another procedure often made use of today is the workbook method. By it, the workbook serves as the motivation for the recitation by being used for reviewing work already studied or for study of advanced material. For this study, that portion of the period in which the workbook is being used will be considered as being based on the workbook procedure.

While it is true that there are other procedures in use today, as well as modifications of the procedures just explained, those defined and explained are the principal ones that lend themselves readily to the nature of the study being made.

DIRECTIONS TO SCORERS

Be in the classroom before the class gets under way. Before scoring a class, make certain that there is nothing unusual to distract the attention of

XUM

the students. Take a position from which every student can be easily observed. Fill all of the blanks at the top of the score sheet.

In the column under teaching procedure, tell exactly what the teacher is doing or having the class do. Whenever possible, indicate which of the following procedures is being used: Class discussion, rapid-fire question-answer, laboratory work, lecture by teacher, use of workbooks, report of student, or demonstration. If in doubt as to the procedure, tell exactly what is being done. When a class discussion is under way, indicate in the procedure column whether the teacher or the student is talking when each count is started. Use *T* for teacher and *S* for student.

Indicate each change in teaching procedure at the time the first count is taken under that procedure. When possible, count the number attending each minute. Use a watch or clock to give the time of each count. If you cannot be certain that a student is attending, do not count him as attending.

An example of a preliminary check upon scorers demonstrates the form used and method of recording. The final scores are the average number attending divided by the number in class to give per cent of attendance.

GROUP CONTROL ATTENTION SCORE—OBSERVER *A*[1]

School Wayne Twp., Subject English II Date April 5, 1939

No. in Class 31 Period begins 2:15 Period ends 3:00

Teaching procedure	Time	Number attending	Teaching procedure	Time	Number attending
Rapid-fire	2:20			2:30	27
questions	21	24		31	30
and	22	23	Discussion *T*	32	30
answers.	23	25	*T*	33	28
	24	23	*T*	34	30
	25	27	*S*	35	28
	26	29	*T*	36	28
	27	28	*T*	37	29
	28	28	*S*	38	29
	29	26	*S*	39	28

[1] This is incomplete, but will serve as an example of the type of recording.

The disagreement in scores or per cents attending was negligible in the above data. However, differences in reported teacher and pupil participation in discussion were more marked. The principal apportioned thirteen minutes to the teacher and eight to the pupils; one student-observer gave twelve to the teacher and nine to pupils and the

other recorded eleven and ten, respectively. Since no observer reported either pupils or teacher to have used seventy-five or a greater per cent of the time, there is no disagreement in the classification of the discussion as general.

TABLE I.—GROUP CONTROL SCORES MADE BY TWO STUDENTS AND A PRINCIPAL DURING AN ENGLISH RECITATION

	Students' scores	Principal's scores
Observer *A* (total score)	88.5	87.9
Observer *B* (total score)	88.9	87.9
Observer *A* (rapid-fire)	87.5	88.1
Observer *B* (rapid-fire)	87.9	88.1
Observer *A* (discussion)	89.2	88.9
Observer *B* (discussion)	89.6	88.9

It should not be concluded from this example that no disagreements were obtained. Many charted observations were discarded during the first semester of the experiment, and three semesters beyond the semester of preliminary procedure were necessary to obtain the data of the study.

At least ten per cent of the observations were failures due to interruptions, outside disturbances, or inside conditions which affected the class attention. Apparatus, board work, and bulletin board material for other classes were constant offenders which led to the discard of completed charts. Naturally, with experience, fewer charts, invalid due to omissions of detail, were received.

COMPARISON OF ATTENTION SCORES UNDER THE DIFFERENT TEACHING PROCEDURES

The following tables present the acceptable data which resulted from approximately one hundred and fifty hours of charted observation in over two hundred classes.

Table II listed the procedures in order of the mean attention (per cent attending) as observed. The ranges show the extremes of the attention scores, and the standard deviations offer a measure of the variability from the mean which considers all scores rather than the two extremes. The number of cases is no indication of the amount of use of the procedures in the high school. Demonstration, laboratories, lectures, and workbook procedures were easy to score due to the con-

tinuity of the procedures. Discussions changed from teacher to pupil or became lectures and were difficult to score because of the notations of these changes which were necessary. This led to the discard of more charts of the types most difficult to score.

TABLE II.—TEACHING PROCEDURES RANKED BY ATTENTION SCORES

A, procedure	*B*, mean	*C*, range of scores	*D*, σ	*E*, number cases
1. Student reports	88.20	77–100	7.73	16
2. Demonstration	87.50	66–100	6.93	69
3. General discussions-*S*	85.03	70– 98	6.49	17
4. Workbook	84.15	70– 98	9.07	22
5. Lecture	84.00	52– 96	9.28	23
6. Rapid-fire	83.00	65– 98	8.16	21
7. General discussion-*T*	82.84	65– 98	9.60	19
8. General discussion	82.30	66– 97	9.94	29
9. Laboratory	80.60	55–100	9.00	50

Reading line 1, column *A*, the student report procedure has a mean attention score, column *B*, of 88.2 per cent with a range of scores, column *C*, from 77 to 100 and a standard deviation, column *D*, of 7.73 for column *E*, the sixteen cases of this procedure.

Table III gave a means of more definite comparisons of the procedures' relative abilities to hold group attention. Consideration of these figures placed student reports and demonstrations in the upper group as superior to all other types in ninety or more cases of one hundred. Student-dominated general discussion must be classed below the upper group but appeared sufficiently superior to the other two types of discussion to be accounted more efficient in obtaining attention. The workbook, lecture, and rapid-fire question methods made up a group without great enough indication of superiority, one over the other, to deserve recognition. However, they showed enough advantage beyond the remaining three to form a group by themselves. Teacher-dominated general discussion and the undominated varied slightly in ability to hold attention and grouped themselves together. While they were not always superior to the last procedure, the laboratory, they were nevertheless likely to show this superiority in obtaining attention in approximately seventy-five of one hundred cases.

These results were borne out in part by the study made by Schaffer[1] who obtained similar results when comparing attention scores under the

[1] Schaffer, Mack: *Op. cit.*, p. 52.

TABLE III.—THE CHANCES IN ONE HUNDRED THAT ONE TEACHING PROCEDURE OBTAINS BETTER ATTENTION THAN ANY OTHER

	Demonstration	General discussion-*S*	Workbook	Lecture	Rapid-fire	General discussion-*T*	General discussion	Laboratory
Student reports	62	88	93	93	94	94	98	100
Demonstration	..	90	94	95	96	96	99	100
General discussion-*S*	..	..	64	65	69	76	86	98
Workbook	..	..	..	52	56	65	76	93
Lecture	..	..	..	..	53	64	74	93
Rapid-fire	..	..	..	..	..	62	73	93
General discussion-*T*	..	..	..	..	..	..	57	78
General discussion	..	..	..	..	..	..	..	77

Reading Table III, following student reports from left to right, in one hundred similar comparisons of observed group attention, student reports would be superior to demonstration in sixty-two cases, to discussion dominated by students in eighty-eight cases, to workbooks in ninety-three cases, to lecture in ninety-three cases, to rapid-fire questioning in ninety-four cases, to discussion dominated by teacher in ninety-four cases, to discussion dominated by neither pupils nor teacher in ninety-eight cases, and to the laboratory procedure in one hundred or every case.

laboratory procedure with those under the demonstration procedure when used in chemistry. In that study the ratio of the mean scores of the two procedures, *i.e.*, the laboratory to the demonstration, was identical to the ratio found in this investigation.

The data on the demonstration procedure provide more conclusive evidence of superiority than do those on the student report procedure because of a larger number of cases. Upon investigation of the original charts, the student reports appear more like the demonstration than the lecture with an informality not apparent in either of the former. Lectures lasted a minimum of five minutes and their average length was ten minutes. The average length of the student report was slightly less than four minutes. Some reports, from their description, were

quite similar to demonstrations in that pictures were presented. Most reports were concerned with current materials in history, social science, and science classes. Evidently the long student report of an explanatory nature had been deleted by the teachers. On the other hand, the lectures were often explanations of more difficult subject-matter which lost attention after the first few minutes of good initial attention. These facts led to the classification of the demonstration and lecture as formal, since both prohibited class participation by discussion for a considerable period; the student report was considered informal, since it appeared similar to student participation in the discussion as found in the classification student-dominated discussion.

In contrast to the initial attention to many lectures is the lack of initial attention shown by many charts of the laboratory procedure with notations to the effect that apparatus and materials were not ready for the students.

IMPLIED QUALIFICATIONS OF GOOD TEACHING PROCEDURE

The procedure which presents the best combination of the following attributes should produce the best attention:

(1) Appropriateness to the learning situation—indicated by the fact that each method maintained satisfactory attention during some of the measured periods. These data imply that pupil participation may not always be measured by how much the pupil does or says as well as by how completely he attends.

(2) Student participation—a conclusion from the data on the three types of discussion which indicated that in the same procedure, student domination of the discussion would be accompanied by better attention in seventy-six of one hundred instances. The fact that student participation would bring inferior attention in twenty-four of one hundred cases designated the appropriateness of teacher domination in these instances.

(3) Thorough previous preparation—suggested by the superiority of the demonstration which requires previous preparation and by the variability in the success of the lecture and laboratory procedures which are definitely dependent upon previous preparation.

(4) Definiteness and clearness of assignment to pupil—demonstrated by the relatively high standing and narrow range of attention scores for the workbook method.

(5) Combined visual and auditory learning—designated by the high attention accompanying the demonstration procedure.

XUM

DERIVING LANGUAGE AND NON-LANGUAGE MEASURES OF INTELLIGENCE FROM THE KUHLMANN-ANDERSON INTELLIGENCE TESTS

GEORGE SPACHE

Psychologist, Brooklyn Friends School, Brooklyn, New York, and Friends Seminary, New York City

Among others, Durrell[1] has demonstrated that reading ability markedly influences the results of intelligence tests. The intelligence of children handicapped in reading is prone to be underestimated when determined by individual or group measures of verbal intelligence. A similar error of underestimation occurs in the testing of bilingual children. As Mahakian[6] has shown, the intelligence of bilinguals is estimated as of a lower level when the test is administered in a language other than that native to the subjects. Because of these tendencies to error, there is need for valid and reliable tests of intelligence which are not unduly influenced by the linguistic abilities of the subjects. An approach is possible through the separate measurement of verbal ability or intelligence, and non-verbal intelligence.

In the writer's use of the Kuhlmann-Anderson intelligence tests, it became apparent that these contain elements of both verbal and non-verbal intelligence. Inspection of the various batteries reveals that they contain, what appear to be, numerous examples of both types of tests. A rapid review of correlations of various Kuhlmann-Anderson batteries with verbal and non-verbal criteria seems to confirm this impression. A group of twenty-eight such correlations has been collected by the writer, but space limitations do not permit their reproduction.

The median correlation of the various Kuhlmann-Anderson batteries with language tests is .679. This is a median correlation with such tests as the Otis S-A, verbal section of the Junior Scholastic Aptitude, Henmon-Nelson, Army Alpha, etc. The median correlation with non-language tests, such as the numerical section of the Junior Scholastic Aptitude, Otis Primary, Pintner-Cunningham, etc., is .687. There is reason to believe that the Kuhlmann-Anderson tests measure both language and non-language abilities. This situation is concealed in the use of a median mental age as now derived from the battery of a particular grade. Certainly the need for separate measures warrants an attempt to identify the language and non-language portions of

XUM

each Kuhlmann-Anderson battery, in the hope that these sections may be utilized to yield differential measurements.

The tests of each grade level may be tentatively identified as measures of language and non-language intelligence by their correlations with recognized criteria of these abilities. For example, the correlations with the respective language and non-language sections of the California Test of Mental Maturity may be used to identify the nature of each Kuhlmann-Anderson test. This procedure assumes that the sections of the California test are valid. The Educational Records Bureau[3] has already given evidence in support of this assumption. They show that the sections of the California test are not identical measures, that the language section is more closely related to verbal intelligence as exhibited by reading ability, and that a distinction between the two sections is warranted. Although more evidence than this is desirable, the present writer has been unable to secure further validating evidence from the publishers of the California test.

In Table I the correlations of the intelligence quotient from each of the Kuhlmann-Anderson tests with the language and non-language IQ's from the California test[8] are given. The Kuhlmann-Anderson tests were given in October of one school year, and the California tests, the following October. Tests 3 to 16 of the Kuhlmann-Anderson were used in the first grade, 8 to 21 in the second grade, 15 to 24 in the third and 18 to 27 in the fourth. This is a somewhat different arrangement

TABLE I.—CORRELATIONS OF IQ'S FROM INDIVIDUAL KUHLMANN-ANDERSON TESTS WITH CALIFORNIA LANGUAGE AND NON-LANGUAGE IQ'S

	Kuhlmann-Anderson Tests												
	3	4	5	6	7	8	9	10	11	12	13	14
California IQ's												
Language	.467	.299	.447	.157	.424	.320	.427	.412	.163	.367	.299	.408
Non-language	.395	.353	.514	.259	.527	.231	.269	.237	.238	.300	.217	.331
N	72	74	75	69	71	143	142	146	149	129	137	103

	15	16	17	18	19	20	21	22	23	24	25	26	27
California IQ's													
Language	.208	.219	.506	.300	.309	.231	.693	.724	.752	.305	.549	.720	.500
Non-language	.412	.269	.402	.263	.206	.208	.506	.490	.451	.534	.408	.570	.463
N	141	208	125	161	165	153	127	111	111	111	43	42	41

than that proposed by the authors of the Kuhlmann-Anderson series.[6] They suggest the use of tests 4–13 in the second semester of the first grade, 8–17 in the second grade, 12–21 in the third, 15–24 in the fourth and 18–27 in the fifth. The arrangement of tests used here is that commonly employed by private schools participating in the programs sponsored by the Educational Records Bureau.[2,3,7] The present writer has previously demonstrated[7] that the use of an enlarged battery in the first and second grades, and of the battery intended for the grade above, in the third and fourth grades, is a sound procedure, on the whole.

Seventeen of the twenty-five tests for which data are available are more closely related to the language section of the California test, and may be tentatively identified as measures of verbal intelligence. The remaining eight tests may be similarly identified as non-verbal measures. The differences between the correlations of each test with the criteria are not statistically significant except in the case of test 23. This is immaterial since the size of the difference is influenced by such details as the number of steps in the scatter-gram, the number of maximum scores in each test, etc. We are concerned with the trend of the differences rather than their exact size.

For the most part, the Kuhlmann-Anderson tests indicated as language measures are those we might have identified by inspection. The identification of tests 8 and 12, the counting of objects or dots, of test 13—a symbol-digit, and 14—a test of completing an *X*-0 sequence as language measures is somewhat unexpected. However, it is possible that these tests may function as measures of verbal facility at this primary level, even though they are frequently considered as non-verbal at higher levels. This is particularly true of the counting tests. The identification of the number tests as language is corroborated by Traxler's evidence.[3] He also found that, contrary to common belief, arithmetical ability was not more closely related to non-verbal intelligence than to verbal. The identification of test 19 might be questioned. It is not clear just why this test of recognition of the parts of common geometric figures should function as a measure of linguistic intelligence, unless the ability tested is analogous to those underlying word analysis. The identification of test 27, an arithmetical division, is in keeping with Traxler's evidence.

There is little reason to question the identification of the eight non-language tests. They include two of pictorial similarities, 7 and 15; two of distinguishing objects according to their properties, 4 and

11; a figure and a design completion, 5 and 24; a memory for designs, 6; and a test of following directions, 16. All of these are of the type generally accepted as measures of non-verbal ability.

The uses of the portions of the Kuhlmann-Anderson identified as language and non-language are obvious. Since the battery for the first grade includes seven tests of each type, both language and non-language median MA's and IQ's may be found, as well as similar measures of ability from the entire test. In the grades above the first, the number of non-language tests is probably too small to warrant the calculation of separate language and non-language measures. Further research is necessary to demonstrate (1) that the separate measures derived from the first grade battery are not identical; (2) that they are measuring different types of intellectual abilities and, therefore, that a distinction is warranted, and (3) that they are valid, or are as closely related to other criteria of language and non-language abilities as other intelligence tests.

One method of demonstrating the validity of the distinction between the language and non-language portions of the Kuhlmann-

TABLE II.—INTERCORRELATIONS OF IQ'S FROM CERTAIN INTELLIGENCE TESTS IN A FIRST-GRADE POPULATION

		Kuhlmann-Anderson non-language	Kuhlmann-Anderson total	Pintner-Cunningham	California total	California language	California non-language	Median
Kuhlmann-Anderson	*r*	.684	.834	.656	.678	.585	.488	.656
Language	*N*	75	75	92	75	75	75	
Kuhlmann-Anderson	*r*	....	.806	.692	.623	.560	.499	.623
Non-language	*N*	....	75	92	75	75	75	
Kuhlmann-Anderson	*r*	....	....	.695	.681	.547	.570	.625
Total battery	*N*	....	....	92	75	75	75	
Pintner-Cunningham	*r*	....	....	....	.589	.545	.574	.622
	N	....	....	....	79	93	93	
California mental	*r*	....	....	....	....	....	.430	.547
Maturity language	*N*							
California mental	*r*	....	....	....	....	....	....	.499
Maturity non-language								
California mental	*r*	....	....	....	....	....	....	.650
Maturity total								

Anderson tests is by determining their relations with each other and similar tests. The median intelligence quotients on the two portions of this test were computed for a number of first-grade children. The correlations of Table II were then computed.

The relationship between the two portions of the Kuhlmann-Anderson tests is about as great as one would expect to find between two somewhat dissimilar intelligence tests used in a single grade range. The relation is slightly, but unreliably, greater than that found between the corresponding portions of the California test. Traxler[2,3] reports correlations between the sections of this latter test of .565, .618, .649, .683, .566, .522 and .586 in various grades. The median of these, .586, is not greatly less than the association between the two sections of the Kuhlmann-Anderson. Since the authors do not give reliability coefficients for the Kuhlmann-Anderson, we do not know whether the relationship between the two sections is less than their reliabilities, or less than that between two random halves of the test. However, it may be concluded that these sections identified as language and non-language are not more identical than other tests of similar nature.

Certain other conclusions are possible from the table. As a rough measure of validity, the median correlation of each test with the others may be used.* This median is given in the last column of the table. It appears that the IQ's from the Kuhlmann-Anderson language and non-language tests are, on the average, as similar to those of other tests, as the IQ's from the Pintner-Cunningham, the entire Kuhlmann-Anderson battery, the two sections or the entire California test. In other words, the two sections of the Kuhlmann-Anderson are apparently as valid measures of intelligence as any of the other tests.

Final evidence of the validity of the distinction between Kuhlmann-Anderson language and non-language IQ's may be secured by the correlations of these with recognized measures of language abilities. If the language IQ's correlate more highly with such a criterion as reading ability, than do the non-language IQ's, it may be presumed that a distinction is warranted. The reading ability measure which served as criterion was the average grade score on three reading tests of the Metropolitan Achievement, Primary I Battery: (1) Word Picture—a test of the ability to associate words and phrases with the

* The correlations of the Kuhlmann-Anderson language and non-language IQ's with those from the entire battery were not used in these averages. These correlations are not, strictly speaking, validity coefficients.

pictures they describe; (2) Word Recognition—identifying words presented auditorily; and (3) Word Meaning—identifying words according to their meanings.

The relationships between the intelligence tests and reading ability are given in Table III. The Kuhlmann-Anderson and Pintner-Cunningham were given in October at the beginning of the first grade, the Metropolitan Achievement in April at the end of the first grade, and the California, the following October at the beginning of the second grade.

TABLE III.—CORRELATIONS BETWEEN INTELLIGENCE AND READING ABILITY IN A FIRST-GRADE POPULATION

	r	*N*	Median IQ	Mean IQ	Median reading	Mean average
Kuhlmann-Anderson language IQ-Metropolitan reading average	.216	84	108.8	109.8	1.82	2.01
Kuhlmann-Anderson non-language IQ-Metropolitan reading average	.110	84	112.8	113.0	1.82	2.01
Kuhlmann-Anderson total battery IQ-Metropolitan reading average	.161	84	109.1	109.8	1.85	1.90
California language IQ-Metropolitan reading average	−.035	98	119.5	120.8	1.90	2.02
California non-language IQ-Metropolitan reading average	.334	98	111.4	111.8	1.90	2.02
California total IQ-Metropolitan reading average	.188	98	113.7	115.0	1.90	2.05
Pintner-Cunningham IQ-Metropolitan reading average	.294	119	118.4	118.5	1.82	1.97

The IQ's from the language section of the Kuhlmann-Anderson battery are not significantly better than the non-language as predictive measures of reading ability at the end of the first grade. Neither of these is significantly better than the entire Kuhlmann-Anderson battery. Thus the primary purpose for a distinction between the two IQ's appears to be defeated.

One of the primary facts found in the table is the low values of all these intelligence tests in the prediction of reading ability at the end of the first grade in this population. As Gates[4] has shown, the size

of the correlation between reading ability and intelligence in the first grade is influenced by the type of the instruction. He claims that the greater the amount of individual instruction, the simpler and more extensive the reading materials, and the greater the provisions for individual differences, the closer will be the association between mental ability and reading success.

In the classes used as a basis for our data, there is, in the opinion of the writer, an average amount of individual instruction and provision for individual differences, of a certain type. The individual instruction consists largely of coaching slower members of the class. This attention to slower students would tend to lower the correlation between mental ability and reading success inasmuch as it tends to raise the reading ability of poorer students to the class median, but does nothing to raise the reading achievement of superior students above the same median.

Moreover, there is another trend which tends to decrease the relationship between mental and reading abilities. The teachers are prone to increase the difficulty of the reading program by lack of continuity in their selection of pre-primers, primers and first readers. It can be demonstrated readily that such a program materially increases the burden of new words and decreases their average repetition. To put it more succinctly, the heavy vocabulary burden prevents the brighter students from achieving the level of which they are potentially capable. On the other hand, the attention to the slower members of the group enables them to achieve at maximum levels. Thus the variability of the group is decreased and the influence of mental capacity lessened. It is also possible that other factors, as maturity, chronological age and absences, are in marked relationship to reading achievement in this population and, hence, are tending to decrease the relationship of intelligence to reading.

It may be that in other first grades in which integrated reading materials are employed, where provision for individual differences at both extremes is made and where other factors play a subordinate rôle to that of intelligence, the relationship between the Kuhlmann-Anderson language IQ's and reading ability would be greater, and approximate the average found for other common tests of intelligence. However, on the basis of the present evidence, there is little reason to believe that there would be marked differences between the Kuhlmann-Anderson language and non-language IQ's in their values as predictive measures of reading ability.

XUM

BIBLIOGRAPHY

1. Durrell, Donald D.: "The Influence of Reading Ability on Intelligence Measures." *Journal Educational Psychology*, Vol. XXIV, September, 1933, pp. 412–416.
2. "1937 Fall Testing Program in Independent Schools." *Educational Records Bulletin, No.* 22. Educational Records Bureau, New York City.
3. "1938 Fall Testing Program in Independent Schools and Supplementary Studies." *Educational Records Bulletin, No.* 26. Educational Records Bureau, New York City.
4. Gates, Arthur I.: "The Necessary Mental Age for Beginning Reading." *Elementary School Journal*, Vol. XXXVII, March, 1937, pp. 497–508.
5. Kuhlmann, Fred. and Anderson, Rose: *Kuhlmann-Anderson Intelligence Tests. Fourth Edition.* Minneapolis, Minn.: Educational Test Bureau, 1933.
6. Mahakian, Chas.: "Measuring Intelligence and Reading Capacity of Spanish-Speaking Children." *Elementary School Journal*, Vol. XXXIX, June, 1939, pp. 760–768.
7. Spache, George: "The Use of the Kuhlmann-Anderson Intelligence Tests in Private Schools." *Journal Educational Psychology*, Vol. XXX, November, 1939, pp. 618–623.
8. Sullivan, Elizabeth T., Clark, Willis W. and Tiegs, Ernest W.: *California Test of Mental Maturity.* Los Angeles, California: California Test Bureau, 1936.

XUM

TWO TEST-RETEST EXPERIMENTS ON KNOWLEDGE OF THE ENGLISH LANGUAGE

AUDRY PRINDLE, R. R. REICHART, AND H. R. LASLETT

Oregon State College, Corvallis, Oregon

The purpose of these studies was the comparison of the knowledge of the English language of college freshmen and college seniors. The first experiment involved one thousand three hundred fifty-nine freshmen who were examined with an English placement test in the Autumn of 1936 and three hundred fifty-three seniors in the same college who were examined with the same test in the Spring of 1940. The second study grew out of the first, and involved the one hundred seventy students who took the examination on both occasions and whose records were compared individually after an interval of forty-four months.

The test used was the Oregon State College English Placement Test. This test is made up of four parts—spelling, grammar and usage, recognition of sentence elements (parts of speech), and punctuation. Each part involves fifty printed sentences, with a possibility of the student making fifty errors. The test is scored in terms of errors made. For that reason, a large score is less satisfactory than a small score on this test. The form of this test has been in use for about nine years, but the contents are worked out anew annually from the same types of material. The principal use of this test is the classification of entering freshmen according to their knowledge of the fundamentals of English composition. Those who score in the highest decile on this test enter English 112, or the second quarter class in a three-quarter freshman sequence. Those whose scores lie in the second to ninth deciles inclusive enter English 111 classes, or the first quarter of the course in freshman composition. Those whose scores lie in the lowest decile enter English *K*, or subfreshman classes. Neither the reliability or the validity of this particular form of this test is available. Since the type and the selection of material in this test are similar to those of standardized tests of similar nature, it is presumed that the reliability and the validity of this test are similar to those of standardized English tests. The 1936 edition of this test was used throughout these experiments.

The scores from the one thousand three hundred fifty-nine freshmen tests were obtained from the records of the results of the use of this test during the annual freshman week in October, 1936. A new

printing of this same test and identical with it was made in 1940 and given to as many seniors as could be reached in May of that year. Generally, these seniors were tested in classes composed almost entirely of seniors. A few groups of seniors who were not in the same classes were gotten together for the tests, also. As the taking of the test by the seniors was voluntary, it is believed that their best efforts were secured. Additional motivational factors were the parts which word-knowledge and analytical comprehension of sentences and paragraphs have come to occupy in the mental and vocational tests which many of these students will take as a part of their employment procedure, and the classes which have been organized for training in this field.

A comparison of the scores of one thousand three hundred fifty-nine freshmen of 1936 and the three hundred fifty-three seniors in 1940 shows that the seniors as a whole were superior to the freshmen as a whole, but the mean scores (in terms of numbers of errors made) of both groups were far from satisfactory when one considers the ele-

TABLE I.—A COMPARISON OF FRESHMAN AND SENIOR CLASSES ON AN ENGLISH PLACEMENT TEST

School	Freshman mean score	Freshman SD	Senior mean score	Senior SD	Gain or loss of senior over freshman mean[1]
Agriculture	90.9	(27.2)	69.2	(20.5)	+21.7
Education	66.6	(27.0)	46.8	(23.8)	+19.8
Engineering	75.0	(27.0)	62.4	(23.5)	+12.6
Forestry	86.5	(25.7)	69.2	(23.1)	+17.3
Home Economics	66.1	(23.6)	52.3	(22.9)	+13.8
Lower Division	70.2	(29.5)[2]			
Pharmacy	61.2	(25.7)	61.2	(24.0)	0
Science	70.3	(25.1)[3]			
Secretarial Science	60.7	(27.3)	40.2	(17.4)	+20.5
Total group	74.3	(27.6)	59.3	(24.4)	+15.0

[1] Actual gains or improvements in scores are shown by plus signs although the improved scores are smaller than the original scores.

[2] The Lower Division consists of freshmen and sophomores only, but many who took the test as freshmen in Lower Division took the test as seniors in other schools.

[3] Because the Science seniors were scattered in many small classes, it was impossible to get them together in worth-while numbers to give them the test under the conditions that existed.

mentary nature of the material of this test. Table I shows the means, standard deviations, and gains or losses of the seniors as a class in comparison with the freshmen.

The most satisfactory mean scores for the freshmen were made by the one hundred eighty freshmen in Secretarial Science and the forty-three freshmen in Pharmacy. The least satisfactory scores were made by the one hundred ninety-four freshmen in Agriculture, the one hundred eighty-four freshmen in Forestry, and the two hundred ninety-three freshmen in Engineering. The standard deviations for the freshmen mean scores were closely similar in size and all were large, showing that each school had about the same proportions of excellent, average, and poor students as far as their high-school preparations in English language were measured by this test. The freshmen as a whole made more than one-third of the possible number of errors. This does not indicate a high degree of proficiency for their training in English in high schools, especially since many of them were trained directly for the taking of this test. The subtest causing these students the most trouble was the one having to do with spelling. The one causing them the least trouble was the one on punctuation.

Among the seniors, the fifty-four in Secretarial Science made considerably more satisfactory scores than those in the other schools and their individual scores showed less deviation from their mean score. The fifty-two seniors in Agriculture and the sixty-three seniors in Forestry who were tested made the least satisfactory scores. The seniors in Agriculture made larger amounts of improvement in their four years than those in the other schools but, contrary to the usual findings, the seniors in Secretarial Science who made the most satisfactory mean score as freshmen made the second largest amount of gain, while those in Forestry who made the second least satisfactory mean score as freshmen were in fourth place among the seven schools in the amount of gain. Similar studies have shown rather uniformly that the schools making the poorest initial showing have made the largest amounts of gain even though they made the poorest final showing as well.

The mean scores of all of the seniors tested were far from satisfactory. In five of the seven schools, the seniors who were tested missed more than one-fourth of the test items. The standard deviations of the senior groups were almost as large as those of the freshman groups, showing that the differences of performance within the groups were large and that there was a great deal of overlapping among the

best performances of the freshmen and the poorest performances of the seniors. This has been shown to be the case in all similar previous studies.

The mean amount of improvement of the seniors over the freshmen was fifteen points. This is twenty-two per cent of the mean score of the freshmen, twenty-five per cent of the mean score of the seniors, or seven and five-tenths of the possible number of errors which might have been made on the test.

In Table II are shown the freshman and senior mean scores, standard deviations, and the gains or losses for the two classes on the subtests which made up the total Oregon State College English Placement Test.

TABLE II.—FRESHMAN AND SENIOR SCORES ON THE SUBTESTS AND ENTIRE TEST

Subtest	Freshman mean	Freshman SD	Senior mean	Senior SD	Mean gain or loss
Spelling	21.7	(9.4)	15.3	(7.8)	+ 6.4
Grammar and usage	16.9	(7.6)	13.7	(6.2)	+ 3.2
Recognition of sentence elements	20.1	(11.5)	18.7	(11.4)	+ 1.4
Punctuation	16.6	(7.0)	11.8	(5.1)	+ 4.8
Entire test	74.3	(27.6)	59.3	(24.4)	+15.0

This table shows that the scores of the freshmen and the seniors were least satisfactory in spelling and in recognition of sentence elements, although the order of these two was reversed during the four-year period. The amounts of improvement in these two show both the largest and the smallest amounts of gain among the four subtests. This is contrary to the usual findings in similar studies, that is, that the largest amounts of gain have been made on the subtests showing the least satisfactory initial scores. Not all previous studies have found this to be true, however.

There may be a certain amount of valid excuse for the unsatisfactory mean score and the large standard deviation of these students on the subtest having to do with recognition of sentence elements. The seniors had, in most cases, not enrolled in any English composition courses since their freshman year and had, probably, not had their attentions called to any points involving sentence elements as such in

other classes excepting those in foreign languages. While the material on sentence elements is elementary in its nature, it is material which is easily forgotten by people not particularly interested in it. The inclusion of this type of material in an English placement test is open to questioning, but the poor performances of these students on the simple material of this subtest is also open to questioning.

The poor showings of these students, as freshmen and as seniors, on the subtests involving spelling, grammar and usage, and punctuation are not as easily excused. The elementary material of these subtests should have been involved in all oral and written recitations in all of the classes of these students over their four years in college, and in high school, as well as in the English classes which they may have had. The mean scores show that the freshmen made errors on one-third of the items of the test, but the seniors made errors on one-fourth of these same items.

The second study is based on the one hundred seventy students who took this 1936 English Placement Test in 1936 and again in 1940. Such a group has been called a "progress group" in other similar studies. The distribution of this group among the schools of the College was roughly proportional to the distribution of the whole student body among the schools. This group, as freshmen, should be slightly superior to the freshmen as a whole because the students who continue their courses for four years are somewhat superior to the majority of those who drop out. In other respects, the group was believed to be a representative group. Table III shows the mean scores and their standard deviations for these one hundred seventy students as freshmen and as seniors.

TABLE III.—TEST-RETEST DATA ON THE ENTIRE ENGLISH PLACEMENT TEST

School	Freshman means	Freshman SD	Senior means	Senior SD	Mean gain or loss	SD
Agriculture	85.4	(20.3)	68.8	(22.0)	+17.9	13.8
Education	49.3	(15.1)	37.5	(21.6)	+12.1	14.2
Engineering	65.0	(26.5)	55.7	(23.5)	+ 7.3	14.4
Forestry	72.5	(17.7)	57.6	(22.4)	+12.5	16.2
Home Economics	52.6	(21.4)	51.5	(22.2)	− .6	13.0
Pharmacy	59.8	(28.2)	57.7	(27.5)	− 4.8	16.2
Secretarial Science	55.4	(19.8)	50.0	(19.4)	+10.6	15.8
Total group	61.9	(26.4)	52.2	(24.0)	+ 8.4	15.8

For this "progress group" of one hundred seventy students, those in the Schools of Agriculture, Forestry, and Engineering made the least satisfactory scores. The largest mean gain among the schools was made by the students in Agriculture even though they made the least satisfactory scores, both as freshmen and as seniors. The largest mean loss was made by the students in Pharmacy, with a loss of 4.8 points. As the number of students in this group was only nine, the loss is subject to considerable statistical doubt. The gains or losses shown in this table were found for individuals, not from means of groups. The students in the School of Education made the most satisfactory scores in both the freshman and senior years, yet made the third largest gain among the schools tested. Their standard deviation was smallest among the freshmen, but not markedly different from the other standard deviations among the seniors. The standard deviation for the students in the School of Pharmacy was so large because the number of students (nine) was so small.

All of the mean scores for the seniors, except that for the students in Agriculture, were superior to the mean score for the freshmen. The mean score for the freshmen in Education is, however, superior to the senior mean for the whole group. The mean score for the freshmen shows that the freshmen missed many more than one-fourth of all of the items. The mean score of the seniors showed that they missed slightly more than one-fourth of the items. Even so, the "progress group" was a little superior to the group composed of all of the freshmen and three hundred fifty-three seniors, both in means and in standard deviations.

Table IV presents the data for the "progress group" of one hundred seventy students by subtests and for the entire test.

TABLE IV.—TEST-RETEST DATA FOR THE SUBTESTS AND THE ENTIRE ENGLISH PLACEMENT TEST

Subtest	Freshman means	Freshman SD	Senior means	Senior SD	Mean gain or loss	SD
Spelling	18.6	(7.4)	13.9	(7.9)	+4.9	5.2
Grammar and usage	14.1	(6.7)	12.6	(6.0)	+ .9	4.6
Recognition of sentence elements	15.5	(9.6)	16.0	(10.7)	−1.3	8.2
Punctuation	13.6	(6.0)	10.4	(4.9)	+3.0	5.7
Entire test	61.9	(26.4)	52.2	(24.0)	+8.4	15.8

This table shows that the largest gains made by this group were on the subtests involving spelling and punctuation, but these gains were quite small. The change from the freshman to the senior mean score was a negligible positive gain. The loss on the recognition of sentence elements was not large, but it was a loss. There was no consistent relationship between the amounts of gain or loss and the rank order positions of the subtests on either the freshman or the senior tests. The data show a very wide spread of knowledge about the elements of English composition in both the freshman and the senior classes. Among the one hundred seventy students who took the test both as seniors and as freshmen, sixty-three of them as seniors, or thirty-seven per cent, earned scores that were less satisfactory than the mean score earned by the one hundred seventy as freshmen; while sixty-two of the freshmen, or thirty-six per cent, earned scores that were superior to the mean score of the one hundred seventy as seniors. When the two classes are compared as classes, the overlapping is only slightly less than when the scores of the same individuals are compared. The results show that the seniors know very little more about the subjects measured by the Oregon State College English Placement Examination than the freshmen do—whether one compares the members of the two classes as groups or the same individuals taking the same test twice with an intervening period of practically four years.

It seems fairly apparent that the problem of training college and university students in the elements of English composition is serious, but that it is little or no more serious in English than in the other subjects which have been measured in a similar manner in other studies. It is to be hoped that the seniors have more functional knowledge than the freshmen even though they appear to have little more factual knowledge. As there is no satisfactory way of measuring the functional knowledge about English of groups of students at the present time, the decision on this point lies in the future. Perhaps the examinations which have been used in the measurement of students' knowledge have not been composed of the material which the students need or want from their courses. This seems, however, to be improbable.

XUM

STUDIES IN THE PSYCHOLOGY OF MEMORIZING PIANO MUSIC. VII. A COMPARISON OF THREE DEGREES OF OVERLEARNING

GRACE RUBIN-RABSON

New York City

No introduction to an overlearning experiment is complete without some reference to Ebbinghaus[3] and his frequently repeated conclusion that each repetition after learning to the threshold has been accomplished has an approximately equal effect upon retention, in that each addition of three learning repetitions means one repetition saved in learning. Cuff[1] not only disproved this fixed relationship, but insisted that overlearning may even hinder retention after twenty-four hours. Jersild,[5] too, observes that the benefit arising from increased frequency does not increase in direct proportion to the number of added repetitions. Krueger[6] found an actual decrease in retention after one day when the learning was carried to the one hundred fifty per cent level, but an increase in the overlearning advantage with a longer time interval; further, as the degree of learning increases from one hundred fifty to two hundred per cent the corresponding increment in retention is usually less than one-third and the proportion bears no consistent relation to the length of time interval. In general, his work shows that overlearning of at least fifty per cent is highly economical for retention over intervals up to twenty-eight days and that the larger the interval the greater is the economy. Greater degrees of overlearning are uneconomical for most intervals.

An experiment in overlearning piano music introduces several novel factors. The material to be learned is intrinsically logical and meaningful; nevertheless, many constituent tonal patterns and hand groupings have already been vastly overlearned and come full-fledged to the new learning. Various structural and developmental devices make for a certain degree of inevitability in some tonal and harmonic progressions so that the very appreciation of their existence is almost a guarantee of their immediate learning. The natural spatial and temporal organization of the musical figures eliminates any need for arbitrary learning associations.

Serious music students are, on the whole, dedicated to a philosophy of long hours of practice and indefinite repetition of study and repertory material. If a firmer line of demarcation were drawn between repe-

titions for learning and finger movements for sheer increase of technical proficiency, a great amount of time might be saved and much more material learned.

The number of trials to the learning point depends in this area, as in most others, on the talent and organizational ability of the learner, and no external device is effective in reducing this number once the coöperation of the learner is assured.[8] With a purposive learning constructed on a foundation of analytical study preliminary to the keyboard trials[7] and a completed product vivid enough to be accurately transcribed, the proportion of overlearning trials for most efficient retention is of vital importance in reducing the student's keyboard drudgery.*

In the experiment antecedent to this one[10] a large amount of overlearning (roughly one hundred per cent) proved no more effective than no overlearning at all. The investigation was concerned primarily with the value of a period of mental rehearsal or "imaginary practice" situated both midway in the keyboard learning and after its completion. With these was contrasted a period of equal length occupied with keyboard trials after the learning point. That the midway study period proved as effective for retention as a greater number of keyboard trials plus the one hundred per cent overlearning may be less an argument against the overlearning than a tribute to the beneficent aspects of the study period. During a second relearning after seven months these differences disappeared, and not even the value of the additional kinaesthesis was in evidence.

That piano music can be learned to the point of perfect transcription without any keyboard practice is incontrovertible, although such an investment of time or any prolongation of such time to include mental overlearning has been found unprofitable.[9] To the experienced musician, the mental image of the material may approach the vividness of actual performance; nonetheless, except the material is short and simple, some inhibition seems to operate during kinaesthetic translation within the first trials.† Nor does mental rehearsal prove as rewarding as other kinds of preparation even after the kinaesthetic learning has been accomplished.[10]

* There is no question of the necessity of sheer "doing" for technical facility. The distinction lies rather in a more economical utilization of the time spent in the repetition of already well-learned material either for more effective technical practice as such or for a greater volume of learned material.

† For talents of artist-virtuoso rank such inhibitions would be present only with increasing length and difficulty.

XUM

THE PROBLEM

The present experiment is concerned only with actual keyboard overlearning. The problem can be stated as follows: Allowing a period for analytical preliminary study, what economies exist in retention if the learning is continued fifty per cent, one hundred per cent, and two hundred per cent past the learning criterion?

THE EXPERIMENT*

Subjects.—With one exception who was older, the nine subjects ranged between twenty-one and twenty-five years of age. All skillful pianists and teachers of piano, they nevertheless represented a relatively wide range in talent and instrumental and theoretical training. This variation can be seen particularly in the average trials to the learning point which ranged from 6.8 to 15.

Experimental Design.—The Latin-Greek square avoids the necessity for equating subjects and materials, and by rotating the several variables involved presents a precise picture of the relative value of the methods. Three of these designs, each followed by three subjects, are set together to form one large design. When each subject performs the experiment three times, this form provides eighty-one learnings, twenty-seven by each of three methods.†

Experimental Materials.—The nine compositions, eight and ten measures in length, were extremely varied in style and in difficulty. Adapted from piano music by Dandrieu, Hook, Balbastre, Rameau, Böhm, Enna, Jarnach, Schmid and Brockway, they represented very old, moderately conventional, and very modern harmonic idioms and provided a fair cross section of actual piano-learning problems.

Procedure.—The nine learnings arranged over four successive days were accomplished individually with only one subject and the experimenter present. After three minutes of analytical study the learning was continued at the keyboard until the criterion of one smooth memorized trial was reached; the trials were continued past this point to the number of half the trials to the learning point for the fifty per cent overlearning; to the same number as the trials to the learning point for the one hundred per cent overlearning, and to twice the number of the learning trials for the two hundred per cent over-

* The experiment was conducted in the studio of the writer in New York City during the Summer of 1940.

† This experimental pattern can be found in [9].

learning. After two weeks and again after seven months the relearning took place without preliminary study. All keyboard trials were done by massed learning,[11] by the "whole" approach[12] and without the benefit of separate-hand learning.[13]

ANALYSIS OF DATA

The learning trials (Table I) to the criterion show only small chance differences. When to these are added the several degrees of overlearning, a wide spread appears between the total trials for *A* (50 per cent), which is 16.56, and *C* (200 per cent) which, at 31.44

TABLE I.—MEANS AND STANDARD DEVIATIONS OF THE LEARNING TRIALS, THE RELEARNING TRIALS AFTER TWO WEEKS AND AFTER SEVEN MONTHS AND THE SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS

Method	Mean	Over-learning	Total trials	S.D.	Difference of means	S.E. differ-ence	Ratio
Learning Trials							
A—50 per cent over-learning	11.04	5.52	16.56	5.54	*A-B*, .82	1.41	0.58
B—100 per cent overlearning	10.22	10.22	20.44	4.60	*B-C*, −.26	1.41	0.17
C—200 per cent overlearning	10.48	20.96	31.44	5.54	*A-C*, .60	1.53	0.39
Relearning Trials after Two Weeks							
A—50 per cent over-learning	6.92	.....	3.52	....	*A-B*, .11	.87	0.13
B—100 per cent overlearning	6.81	.....	2.72	....	*B-C*, .77	.66	1.17
C—200 per cent overlearning	6.04	.....	2.00	....	*A-C*, .88	.79	1.11
Relearning Trials after Seven Months							
A—50 per cent over-learning	6.55	.....	1.77	....	*A-B*, .51	.46	1.11
B—100 per cent overlearning	6.04	.....	1.64	....	*B-C*, .22	.41	0.54
C—200 per cent overlearning	5.82	.....	1.37	....	*A-C*, .73	.43	1.69

trials, is nearly double the trials for *A*. *B* with 20.44 trials lies closer to *A* but is about two-thirds of the number for *C*.

The summary of the relearning trials after two weeks (Table I) presents an unusual picture. Not only do the means show a difference of less than one trial in spite of the large differences in the total learning trials, but the standard deviations are relatively small. Since the overlearning was proportional to the learning, the slower learners accumulated an absolutely much greater total number of learning trials that the more gifted and reduced their relearning trials relatively more than did the speedier ones. The speediest learner showed an average of six and eight-tenth learning trials and five relearning trials, while the slowest, with fifteen learning trials averaged only seven and four-tenth relearning trials. This situation reduced the ordinarily high learning-relearning correlation to 0.35. (In another experiment, using the same subjects,[10] this index stood at 0.71.) In spite of this, when the retention scores of the three most gifted and the three least gifted subjects are compared, there is some evidence that the most gifted profited more from the increased learning trials than did the least gifted.

After seven months the relearning (Table I) trials indicate little loss. The means lie close together as before with a difference of less than one trial. The larger standard deviation may be proof that the more able learners retained relatively more than the less able. Means and standard deviations in both relearnings are in every case progressively smaller from *A* to *C*, but none of the critical ratios approaches a significant level, nor could these small savings justify such a large investment of learning trials.

Table II, an analysis of the compositions ranked according to difficulty, is presented as a corollary to the main data for the three methods. For each composition the required trials for three methods were averaged. This figure appears in column 1. In column 2 appears the number of overlearning trials, secured by averaging the overlearning by three methods, or one hundred sixteen per cent. In column 3 are the total trials of learning and overlearning. The means of the trials of the two relearnings are presented with a percentage index of the savings resulting from the original investment.

The three most difficult compositions were all in the modern idiom. One presented a unique confusion, comical in one sense in view of the experience of the subjects, significant in another for the improvement of music typography. A simple pattern written on the leger lines

high above the treble staff defied organization but would have been immediately comprehensible had it been written on the staff with the customary notation for transposition of an octave.

TABLE II.—COMPOSITIONS RANKED ACCORDING TO DIFFICULTY SHOWING THE AVERAGE LEARNING, OVERLEARNING AND TOTAL TRIALS; AVERAGE RELEARNING TRIALS AFTER TWO WEEKS AND AFTER SEVEN MONTHS AND THE PERCENTAGES OF SAVINGS IN RETENTION

	Learning trials			Relearning trials after two weeks			Relearning trials after seven months		
	Means	116 per cent over learning	Total trials	Means	Per cent of over-learning	Per cent of total	Means	Per cent of learning	Per cent of total
1	6.5	7.6	14.1	5.4	83	38	5.1	78	36
2	6.8	7.9	14.7	5.2	73	35	5.5	81	30
3	7.9	9.2	17.1	5.0	63	29	4.8	61	28
4	9.5	11.1	20.6	6.2	65	30	7.2	76	35
5	9.8	11.4	21.2	6.1	62	29	6.2	63	24
6	11.5	13.4	24.9	7.7	67	31	7.4	64	29
7	12.4	14.5	26.9	6.8	54	25	6.8	55	25
8	13.8	16.1	29.9	6.4	46	21	7.4	53	25
9	16.9	19.7	36.6	6.3	37	17	8.7	52	24

TABLE III.—ANALYSIS OF THE VARIANCE OF THE LEARNING TRIALS AND THE RELEARNING TRIALS AFTER TWO WEEKS, SHOWING THE PERCENTAGE OF THE TOTAL DUE TO EACH VARIABLE, AND A COMPARISON OF THE RATIO OF THE VARIABLE AND THE EXPERIMENTAL ERROR WITH FISHER'S ONE PER CENT VALUE

Variable	Fisher's one per cent value	Learning trials		Relearning trials	
		Ratio	Per cent	Ratio	Per cent
Subjects	2.82	5.4	28	2.8	24
Methods	4.98	0.4	2	2.3	19
Compositions	2.82	8.5	44	3.6	31
Order	4.98	4.0	21	2.0	18
Error	....	...	5	...	8

The analysis of the variance (Table III) supports the insignificance of the methods in relearning. The ratio of this variable with the experimental error is well below Fisher's[4] one per cent value, that

value which might be exceeded in random sampling from a homogeneous population once in a hundred times.

There is ample evidence that the one hundred and two hundred per cent overlearnings are no more effective for retention than fifty per cent overlearning. Whether fifty per cent is more effective than no overlearning can not be concluded directly from the results of this situation; but at the risk of some illogic, a trend may be deduced from another situation[10] in which a one hundred per cent overlearning proved no more economical than no overlearning.

In a learning area which, superficially, appears to be predominantly kinaesthetic this conclusion may seem suspect. It seems reasonable to believe that a muscle skill is improved by "doing," though even this assumption is challenged by Cuff,[2] who says that the belief that "the frequent passage of a neural impulse somehow leads to a fixation of a habit is unsubstantiated." Whatever may be said for repetition in overlearning, it is inescapable that repetitions to reach the learning point do not in and of themselves produce dependable learning.[14] This statement of Thorndike's as well as Welborn's observation[15] that repetition is not of major importance in logical memory lend further credibility to the results of the experiment already described.[10]

SUMMARY AND CONCLUSIONS

Much more than another challenge to the "law of use" is involved in a study of overlearning in a music-memory area. In the voluntary acceptance by piano students of the inevitability of long hours of practice, an enormous amount of time is wasted not only in inefficient learning techniques but in constant repetition of already learned material. Unquestionably sheer "doing" is requisite to the perfection of a comprehensive finger technique; nevertheless, a distinction must be made between time spent in actual learning which may or may not increase dexterity concomitant with the learning, and an emphasis on hand manoeuvers which will provide the maximum technical growth in a given amount of practice.

To evaluate proportional degrees of overlearning past the criterion of smooth memorized performance, the present experiment added fifty, one hundred and two hundred per cent of the learning trials to the finished product. After two weeks, although all three degrees of overlearning achieved large proportions of retention when compared with the learning trials, none was more effective than another. The

two hundred per cent learning involved roughly twice the trials of the fifty per cent procedure, yet showed no superiority in relearning. This being so, it is not surprising that the one hundred per cent overlearning, which provided about one-fourth more trials, was no more effective than the fifty per cent overlearning. How very secondary to the intellectual aspects of the logical learning is the kinaesthetic factor, is indicated by the results of a second relearning after seven months. No trace appeared of any lingering tonal or hand patterns due to the greater learning investment. How good a learning is insured by an emphasis on pre-keyboard analysis plus an immediate memorizing under the intensive conditions of the experimental situation is shown by almost no differences in the means after seven months and after two weeks.

Degrees of overlearning greater than fifty per cent are apparently ineffective for retention. In the present situation no comparison was made of fifty per cent and zero overlearning, but if it is not illogical to deduce a conclusion from another experiment using the same subjects in which a one hundred per cent overlearning proved no more effective than no overlearning at all, then it might be recommended that the most economical learning procedure is one which carries an intensive logical learning to the criterion of smooth memorized performance and saves the extra trials for subsequent practice periods to restore the learning to this same level.

BIBLIOGRAPHY

(1) Cuff, N. B.: *The Relation of Overlearning to Retention.* Geo. Peabody Coll. for Teachers, Contr. to Educ., No. 43, 1927, pp. 34.
(2) Cuff, N. B.: "The Law of Use." *J. Educ. Psych.*, Vol. xx, 1929, pp. 438–447.
(3) Ebbinghaus, H.: "Ueber das Gedächtnis." *Untersuchungen zur Experimentellen Psychologie.* Duncker und Homblot, Leipzig, 1885.
(4) Fisher, R. A.: *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh, 1934.
(5) Jersild, A.: "Primacy, Recency, Frequency and Vividness." *J. Exp. Psych.*, Vol. xii, 1929, pp. 58–70.
(6) Krueger, W. C.: "The Effect of Overlearning on Retention." *J. Exp. Psych.*, Vol. xii, 1929, pp. 71–78.
(7) Rubin-Rabson, G.: "The Influence of Analytical Pre-study in Memorizing Piano Music." Archives of Psych., No. 220, 1937.
(8) Rubin-Rabson, G.: "Studies in the Psychology of Memorizing Piano Music IV. The Effect of Incentive." *J. Educ. Psych.*, Vol. xxxii, 1941, pp. 45–54.
(9) Rubin-Rabson, G.: "Studies in the Psychology of Memorizing Piano Music. V. A Comparison of Pre-study Periods of Varied Length." *J. Educ. Psych.*, Vol. xxxii, 1941, pp. 101–112.

XUM

(10) Rubin-Rabson, G.: "Studies in the Psychology of Memorizing Piano Music. VI. A Comparison of Two Forms of Mental Rehearsal and Keyboard Overlearning." *J. Educ. Psych.*, Vol. XXXII, pp. 593–602.

(11) Rubin-Rabson, G.: "Studies in the Psychology of Memorizing Piano Music. II. A Comparison of Massed and Distributed Learning." *J. Educ. Psych.*, Vol. XXXI, 1940, pp. 270–284.

(12) Rubin-Rabson, G.: "Studies in the Psychology of Memorizing Piano Music. III. A Comparison of the Whole and the Part Approach." *J. Educ. Psych.*, Vol. XXXI, 1940, pp. 460–476.

(13) Rubin-Rabson, G.: "Studies in the Psychology of Memorizing Piano Music. I. A Comparison of the Unilateral and the Coordinated Approaches." *J. Educ. Psych.*, Vol. XXX, 1939, pp. 321–345.

(14) Thorndike, E. L.: *Human Learning.* Century Co., New York, 1931.

(15) Welborn, E. L.: "Logical Learning and Retention. A General Review of Experiments with Meaningful Verbal Materials." *Psych. Bull.*, Vol. XXXIV, 1937, pp. 1–20.

XUM

THE CONCEPT OF "HIGHEST BINET ATTAINMENT" APPLIED TO UNSELECTED CLINICAL CASES

EDITH A. DAVIS

Consulting Psychologist, Idaho Child Welfare Services, Boise, Idaho

The clinical psychologist whose findings institutionalized a moron, or denied him vocational training or surgical correction under a Federal program, is sometimes subject to doubt when reviewing cases. Bijou's[2] distinction between real and pseudo-feeblemindedness, with a fool-proof short cut for making the discrimination, has therefore an immediate appeal. Like Speer[8] and doubtless many other workers, I was impelled to apply the new method to an accumulation of data, to ascertain whether its use would have affected the final disposition of any cases.

Clinical psychology in Idaho is a new departure. Until the present program was initiated in 1939[3] commitments were usually made without psychometric examination, and neither the State Department of Education nor that of Public Welfare employed a psychologist. An occasional child whose Binet IQ was afterwards found to be well above 100 was committed as feebleminded, while some educable children seem to have been excluded from the public schools. It is imperative that no such mistakes be made under psychological auspices, if the program is to develop. Bijou's suggestion is therefore especially pertinent to the situation.

Since the policy in opening this new field has been to encourage the interest of all agencies, referrals have varied over the entire range of ability, with complications involving every type of handicap, neurosis, and personality problem. Brilliant youngsters were tested as much to keep the work from being associated with mental deficiency and poor school work as from a genuine need of interpretation to family and school, in preparation for educational and vocational planning. One or two were deliberately chosen for demonstration tests.

Available for this analysis were three hundred sixty-seven Revised Stanford-Binet examinations, Form *L*. Chronological ages ranged from below five years to the adult level, and intelligence quotients from 20 to 155. Many special handicaps were involved, since the reason for referral was frequently to determine educability, the feasibility of speech correction, and so on. There were two hundred twenty-three males and one hundred forty-four females. Table I

XUM

gives the distribution by age, sex, and ability grouping, using the usual classifications. Many factors peculiar to circumstances involved in particular cases undoubtedly affected the fluctuations in sex ratio for different ages and ability groupings, but no conclusions on the basis of this break-down are warranted. Bijou's subjects

TABLE I.—DISTRIBUTION OF CASES BY CHRONOLOGICAL AGE, SEX, AND ABILITY GROUPING

Age in months	Very superior (IQ above 130)		Superior (IQ 110–129)		Average (IQ 90–109)		Dull (IQ 80–89)		Borderline (IQ 70–79)		Moron (IQ 50–69)		Imbecile (IQ 20–49)	
	M	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*
192—	1	0	4	2	7	4	2	8	1	5	5	6	1	2
180–191	0	0	4	0	3	1	6	1	3	3	3	2	2	2
168–179	0	0	1	3	7	4	4	1	1	1	4	2	0	1
156–167	2	2	3	1	7	0	2	3	2	6	3	3	2	3
144–155	1	0	1	0	6	2	1	1	2	1	4	2	2	3
132–143	2	0	1	1	5	1	2	3	3	2	3	3	1	4
120–131	1	0	2	0	5	1	5	0	6	2	3	2	0	0
108–119	1	0	3	1	3	3	2	2	3	6	4	0	2	0
96–107	2	0	0	0	10	2	4	1	3	3	6	3	0	1
84– 95	0	0	0	0	3	4	7	2	4	6	6	1	3	0
72– 83	0	1	1	1	2	5	3	3	2	1	1	1	2	0
60– 71	0	0	0	0	2	1	1	0	0	0	2	0	2	0
—59	0	0	0	0	0	0	1	1	1	1	1	0	0	0
Total.....	10	3	20	9	60	28	40	26	31	37	45	25	17	16

were free from speech defects, hearing loss, and language handicaps. Speer does not state whether his "one hundred eight children whose IQ's are below 70" were thus selected, but one assumes not. Bijou's age range was from seven to sixty-three years. Speer does not give the chronological age, nor does he distinguish sex or race. Neither author states whether the recorded "Upper Binet Attainment" allowed for the occurrence of successes above the level of total failure, or of failures below the level of the basal age, or, in other words, the now frequently reported finding of more than one basal age.[1,7]

My own practice is to continue the test until I am convinced that no more items will be passed. Conversely, if a child passes all items at six years, yet fails to repeat five digits at seven years, I do not

credit him with the items below seventy-two months until I have made sure that he is able to repeat four digits. Similarly, if a child fails the paper cutting test for the nine- as well as for the thirteen-year level, even though he succeeds with all the twelve-year tasks, he is not credited with the ability to reproduce designs from memory until he has actually passed the test. The discovery of this gross lack of visual imagery in a boy of fifteen, IQ 95, who wanted to study medicine yet was failing algebra and Latin, was very valuable in working out vocational and academic plans that were convincing to the school faculty, yet acceptable to the boy and his family. On the other hand, if an occasional dull youngster can pass the paper cutting test at the third superior adult level, the significance of the fact for vocational guidance is out of all proportion to the few IQ points added to the score. This ability should be unearthed, although the resulting picture of "scatter" may be disconcerting. Thus a girl, IQ 78, CA one hundred sixty-nine months, basal MA eighty-four months, by passing this item earned a new IQ according to the "highest Binet attainment" method of 148, with scatter over one hundred seventy-two months, an obvious absurdity.

This procedure of course increases the total range of success and failure and frequently produces two basal ages. The increase in IQ is seldom more than two or three points, yet the insight into the child's grasp of practical affairs, methods of work, or special abilities may be very illuminating. This is particularly important in a beginning program, where decisions must often be made in emergencies, yet the future prestige of clinical work demands that error be held to a minimum.

Factual reports on scatter for the 1937 Revised Stanford are of interest, particularly as some workers apparently are still limiting the extent of the examination to the range between the established basal and the level at which all tests are failed. My material has been analyzed for scatter as well as for gain in IQ when measured by highest attainment scoring, and group comparisons have been made by intelligence level, chronological age, sex, and special problems for which referred.

"Scatter" for the purposes of this investigation is defined as the number of months of mental age between the lowest basal and the highest success, both inclusive. Thus, a child whose basal age is forty-two months and whose highest success is the pictured absurdities at eighty-four months has a highest Binet attainment MA of seventy-

XUM

four months, and a scatter of thirty-eight months. This procedure was adopted empirically, although a slightly different system would have been equally justifiable.

Scatter was slightly but not reliably greater for males than for females. The means were 65.7 and 62.0 months of mental age, with a critical ratio between means of 1.09. The mean increase in IQ using the highest Binet attainment method was 19.4 points for males and 17.9 for females. Tables II and III show that older subjects and

TABLE II.—MEAN INCREASE IN IQ USING "HIGHEST BINET ATTAINMENT" AND MEAN MONTHS OF SCATTER AT SUCCESSIVE AGE LEVELS

Age in months	Increase in IQ			Months of scatter		
	Males	Females	Both	Males	Females	Both
192—	19.9	17.8	18.7	94.3	81.0	86.8
180–191	17.8	17.2	17.7	78.1	70.9	76.0
168–179	20.9	22.9	21.7	85.8	87.0	86.3
156–167	18.1	19.8	18.9	77.4	74.1	75.9
144–155	26.3	14.0	22.1	81.9	59.1	74.0
132–143	23.1	17.0	20.3	80.6	54.1	68.6
120–131	23.8	18.8	22.8	72.4	54.8	69.1
108–119	22.2	26.8	24.0	60.0	66.5	62.6
96–107	19.2	12.9	17.3	49.2	37.4	45.8
84– 95	11.7	14.1	13.7	29.2	37.2	32.0
72– 83	13.5	15.2	14.4	31.8	35.6	33.8
60– 71	14.4	11.0	14.0	25.4	38.0	27.0
—59	15.3	15.5	15.4	26.7	19.0	23.6
Mean....	19.4	17.9	18.8	65.7	62.0	64.2

superior subjects tend to make the greatest gain in IQ when tests are rescored in this way. The amount of scatter is obviously greatest for these subjects. It appears, then, that when Bijou's suggestion is applied to all types of cases, rather than to the feebleminded alone, the IQ is raised about 19 points, but the relative position of the individual is rarely changed.

Considering now the specific problems for which subjects were referred, we find that the greatest increase in IQ by the new method of scoring is made by those with reading difficulty. For thirty-six cases the mean gain is 26.38 points. Those referred for personality difficulties gained 22.00 points, those needing vocational or educational

XUM

guidance, 19.3 points, the delinquents 18.19 points, and the speech defectives 16.84 points. It seems probable that in the case of reading disability the amount of gain does reflect inadequacy in the regular method of scoring. Critical ratios between mean gains for reading cases and for other types of problems are 3.22 for speech, 2.93 for

TABLE III.—MEAN INCREASE IN IQ USING "HIGHEST BINET ATTAINMENT" AND MEAN MONTHS OF SCATTER BY ABILITY GROUPINGS

Ability	Increase in IQ	Months of scatter
Very superior	22.4	94.9
Superior	19.6	80.3
Average	22.1	71.8
Dull	19.7	66.5
Borderline	19.8	62.1
Moron	15.7	52.5
Imbecile	10.7	34.8

delinquency, and 3.39 for guidance cases. Many cases referred for personality quirks proved also to have reading difficulty, and some were included in both lists. For this reason the critical ratio between these two types was not calculated. Between gain in speech cases and personality cases, the critical ratio was 1.68. The amount of scatter also was significantly greater for the reading than for the speech

TABLE IV.—CRITICAL RATIOS BETWEEN MEAN GAIN FOR READING CASES AND FOR OTHER TYPES OF PROBLEMS WHEN TESTS ARE RESCORED BY "HIGHEST BINET ATTAINMENT" METHOD

Problem	Number of cases	Mean increase in IQ	Critical ratio
Reading	36	26.4	
Speech	42	16.8	3.22
Delinquency	26	18.2	2.93
Guidance	45	19.3	3.39
Personality	44	22.0	

cases. The means were eighty-one and forty-six months, respectively, with a critical ratio of 5.62. However, most of the reading cases were more than eight years old, while many speech defectives were referred at the age of six or seven, or even younger. Table IV summarizes these findings.

XUM

Even with the reading cases there seems to be no practical advantage in using Bijou's procedure. The Spearman rank-order correlation between original IQ and Highest Binet Attainment IQ was .81 PE .04, for the reading cases. The IQ of the brightest child was raised from 119 to 146, second in rank, by the new method. His Arthur Performance IQ was also 119. It is true that his teachers had suspected subnormality because he could not read, but they accepted the fact of his superiority, and coöperated in plans for special instruction. The boy who ranked lowest by the regular scoring was also lowest by the new method. His original IQ of 57 became an IQ of 66, and his Arthur IQ was 69. He is clearly subnormal, with complications of reading disability and maternal over-protection. Almost always the teachers of these children had recognized those who might respond to specialized methods, on the basis of their accomplishment along other lines. In no instance does the new technique indicate a change in the original recommendations.

For sixty of the three hundred sixty-seven cases Arthur Point Scale IQ's were available. For thirty of these the Arthur IQ was higher than the Stanford IQ by an average of 15.4 points; for twenty-four cases it was lower (mean 9.36 points); and for six the IQ was the same. The mean increase in Stanford IQ by highest attainment scoring was 21.97 points for those whose Arthur IQ was higher; 26.0 points for those whose Arthur IQ was lower; and 18.67 points for those whose Arthur IQ was the same as the Stanford. This does not indicate that the new technique can be relied upon to correct the Stanford, which is said to penalize those subjects who tend to make better scores on performance than on verbal tests.

Since Bijou's method makes the erroneous assumption that success at a given level presupposes success on all items below that level, and ignores the fact that differences in abilities occur within the individual, it must not be adopted in clinical practice. Its effect is to raise the obtained IQ in almost every case, but affects the relative standing very little. Its use would seldom if ever change the decision of an experienced clinician, whose recommendations for commitment are based on the modern flexible interpretation of feeblemindedness, as a composite of general intelligence, specific disabilities, social and emotional maturity, and personal adjustment. These conclusions, following the analysis of cases varying over the entire range of ability, are essentially the same as those of Speer,[5] who considered only feebleminded subjects, with a control group of superior ones.

XUM

SUMMARY

(1) The mean increase in IQ for three hundred sixty-seven unselected clinical cases when tests were rescored by the "highest Binet attainment" method was about 19 points.

(2) The greatest increase was made by the older and brighter subjects. It was slightly more for boys than for girls.

(3) Subjects with reading disability showed the greatest gain, but even for these cases the relative standing was little affected by the re-computation of IQ.

(4) Subjects whose Arthur Performance IQ was lower than the Stanford IQ made a greater gain by rescoring than those whose performance IQ was higher than the Stanford.

(5) Bijou's technique is not useful in clinical procedure.

REFERENCES

1. Berger, A., and Speevack, M.: "An analysis of the range of testing and scattering among retarded children on form *L* of the revised Stanford Binet." *J. educ. Psychol.*, vol. XXXI, 1940, pp. 39–44.
2. Bijou, S. W.: "The problem of pseudo-feeblemindedness." *J. educ. Psychol.*, vol. XXX, 1939, pp. 519–526.
3. Davis, E. A.: "Clinical psychology in the Idaho Program of Child Welfare." Accepted for publication by *The Child*.
4. Fernald, M. R.: "A comparative study of scatter on the original and the revised Stanford-Binet scales by the use of retest data." *Psych. Bull.*, vol. XXXVII, 1940, p. 510.
5. Gillette, A. L.: "The accumulation of some data from clinical material on the revised Binet. Form *L*." *Psych. Bull.*, vol. XXXVII, 1940, p. 512.
6. Lorr, M., and Meister, R. K.: "The concept of scatter in the light of mental test theory." *Psych. Bull.*, vol. XXXVII, 1940, p. 513.
7. Mitchell, M. B.: "Irregularities of university students on the revised Stanford-Binet." *Psych. Bull.*, vol. XXXVII, 1940, p. 511.
8. Speer, G. S.: "The problem of pseudo-feeblemindedness: A reply." *J. educ. Psych.*, vol. XXXI, 1940, pp. 693–698.
9. Wile, I. S., and Davis, R : "A study of the basal age with reference to its meaning for school adjustment." *Amer. J. Orthopsychiat.*, vol. VII, 1937, pp. 441–455.

XUM

BOOK REVIEWS

OSCAR K. BUROS. *The Second Yearbook of Research and Statistical Methodology Books and Reviews.* Highland Park, N. J., The Gryphon Press. 1941, pp. 383.

The task of reviewing a book of reviews, such as this yearbook, is a peculiar assignment, especially when one realizes that what is being written is apt to be included in the next yearbook. This volume contains one thousand six hundred fifty-two review excerpts pertaining to three hundred fifty-nine books on statistical and research methodology which have appeared in English since 1932. These reviews have been gleaned from two hundred eighty-three different journals.

It is stated in the preface that the value of reviews is greatly diminished for two reasons: Incompetent reviewers and inaccessibility of reviews. This volume is obviously aimed at overcoming the second difficulty, although the compiler hopes that an eventual outcome will be an improvement in the quality of reviews. We are in marked sympathy with this hope, but we are not quite sure that such a compilation will hasten its realization.

In the preface one finds eleven specific objectives of this yearbook. The first three can easily be telescoped into one: To aid teachers and students in the evaluation of books. Laudable as this may be, it leads to an interesting dilemma: One must depend upon inadequate reviews or one must somehow be able to sense which reviewers are really competent. If the compiler could devise some scheme for appraising the reviews (and reviewers), then publish only those reviews which are actually scholarly and list the remaining as "also found," two purposes would be served: First, the reader would place more faith in the reproduced reviews, and second, the "also wrote a review" would thereby be informed that some super reviewer of reviews had found his review of little value. This might lead editors to be more careful in choosing reviewers who in turn would "take their responsibilities more seriously."

The ultimate value of such a volume as this can only be determined on the basis of its actual usefulness. The person who desires to see what reviewers say about a given book on statistical and research methodology will find this volume a real convenience. The general style of presentation is commendable, and the indexing is such as to facilitate the utility of the volume.

QUINN MCNEMAR.
Social Science Research Council.

XUM

CARTER, ALEXANDER. *How to Locate Educational Information and Data.* New York: Bureau of Publications, Teachers College, Columbia University, 1941, pp. 439.

Graduate students in education probably spend too much time in the library reading and too little time: (*a*) Observing children, and (*b*) thinking about what they have read and observed. While the great respect that the author of the volume being reviewed has for the written page would probably cause him to dissent from this contention, his new edition of the 1935 manual of procedures for locating pay-dirt in the library will tend to alleviate the condition somewhat. The reason is that Alexander's suggestions undoubtedly enable the student who is bent on reading about his problems to save a great deal of the time he usually wastes fumbling with the sources.

The educationist who was familiar with the earlier edition will find this revision considerably enlarged and improved. About one hundred fifty pages have been added, certain chapters expanded and a few entirely new ones introduced. The discussion of the use of the library card catalogue and the Education Index is much more complete in the present book, and a new section appears on "Making a Bibliography." The chapters dealing with the location of biographical information, names and addresses were expanded and four entirely new chapters or appendices included under these titles: Audio-Visual Aids, Maps and Graphs, The Technique of Library Searching, and Borrowing from Other Libraries.

The pedagogical effectiveness of the new edition would seem to be promising. The reviewer did not have at hand the volume of learning experiences, but the emphasis throughout the text is on the importance of learning about locating educational data by trying to locate them. The book's organization is something of a compromise between, or combination of, a description of the nature and use of various library sources such as reference books, card catalogues and the Education Index, and a statement of where to go when concerned with locating information about specific educational problems such as the legal aspects of education or audio-visual aids.

The only mistake that Alexander seems to have made, and it is probably not a serious one from the point of view of graduate students, is his assumption that most of his readers will want to use the library for all purposes. He may be right, but the reviewer was most interested in locating hints about how to get at material dealing with

XUM

specific topics such as educational methods, child psychology, public school administration and educational finance. He was disappointed because the specific topics considered are too few. Anyone who persists in reading chapter after chapter learns a great deal about the value and use of the various materials in the library, but an alternative organization in terms of educational problems, or even more discussion of the best places to go to get answers to specific questions related to large areas, would have helped. As it is the volume is of much greater value as a text than as a handbook.

In the light of this observation, the book would seem to be best adapted to the needs of educational librarians or graduate students who must work up many bibliographies on many topics during the course of the year and who must consequently take time out to get an overview of the resources of the entire library. Probably some public school officials will also come in this group. It would be interesting to know the extent to which the earlier volume was used outside of institutions of higher learning.

The author's style is salty and readable. His pride in his effort is excusable in view of its general excellence. The preface reads: "The present edition of this book and its accompanying experiences differ from the pioneer, or 1935, edition as a 1941 streamlined automobile differs from the 1935 model of the same make."

STEPHEN M. COREY.

University of Chicago.

C. GILBERT WRENN AND D. L. HARLEY. *Time on Their Hands.* Washington, D. C.; American Council on Education, 1941, pp. 256.

America needs and expects an enthusiastic response from its youth for maintaining our defenses. This it cannot get unless sufficient attention is given to the pressing needs of youth. One of the pressing needs is healthful recreation. Basic to a well-planned program aimed to cope with this problem is knowledge concerning the present status of the recreational field. This volume on the leisure time of youth, prepared for the American Youth Commission, gives precisely the information needed by recreational workers and social workers who are concerned with the problem.

Fairly adequate discussions of agencies, public as well as private, which do work in recreation, are found in the book. Also, included are fairly adequate discussions of the nature of the work and the extent

to which they do and do not satisfy the needs of American youth. American youth spend time listening to the radio, seeing movies, and reading. What they get out of it, what can be done about improving the nature of such recreational facilities, is discussed in this volume. Also discussed and to some measure evaluated are the functions of schools and libraries and similar publicly supported agencies; the services of Y.M.C.A.'s, Boy Scouts, settlement houses, and other privately supported community agencies; the work done in the Civilian Conservation Corps camps, the N.Y.A., and some of the relevant W.P.A. projects.

The fundamental thinking behind the author's discussion of recreation is summarized in three principles: (1) The determining factor in recreational planning should be the needs of young people themselves, (2) the greatest possible use should be made of existing facilities and the fullest measure of coördination effected among them, (3) a broad scale method of recreational planning is urgently required.

Psychologists who will search for a treatment of the compensatory function of play or similar concepts will not find them in this volume. In epitome the recreational problem of youth as given by the authors includes such factors as sincere but insufficient efforts of the school, lack of understanding at home, and non-school opportunities restricted largely to commercial places of entertainment. Which is to say that what the authors are concerned with is what clinical sociologists like Clifford Shaw have been emphasizing for many years in their studies of delinquents. Because the book deals with realistic considerations and gives an adequate description of the general field of recreation in community, state, and nation, it should be of help to social planning councils and particularly recreational workers in such councils as well as to other recreational workers, particularly those who do public rather than private work. Worthy of consideration for social planning councils are the major objectives and recommendations that the authors give in the last chapter of the book.

H. Meltzer.
Psychological Service Center, St. Louis, Missouri.

E. A. Betts and A. S. Austin. *Visual Problems of School Children.* Chicago: The Professional Press, 1941, pp. 80.

The reports in this collection of papers describe the data obtained in a comprehensive inventory of seeing and visual status for a fifth-grade population of one hundred twenty-six childern. In addition

to measures of reading proficiency and intelligence, data were secured on the Betts Visual Sensation and Perception Tests. Also a regular analytical visual routine and cycloplegic examination were made by an ophthalmologist. Although the visual findings were compared with reading achievement, the main emphasis was on describing "for the first time a technique of examination based upon a psycho-physiological concept of seeing." Seeing is considered to be functional and is to be contrasted with the inadequate concept of vision which emphasizes structure as a factor. Although both structure and function should be considered, emphasis should be placed upon function. Patterns of seeing are said to be learned. Thus faulty patterns of seeing "may be corrected not only by changing the stimulus by means of glasses but also by re-orientation of the individual through re-education." Most psychologists will approve of a functional, dynamic emphasis when considering problems of seeing. But the authors have not made clear what they mean by the "patterns of seeing which are learned." It would be desirable, therefore, to have a more complete treatment of this psychophysiological approach to visual problem.

Certain other contributions may be cited: (1) The insistence that near point and binocular tests be made in addition to the far point and monocular tests is adequately supported by experimental evidence and argument. (2) No significant relationship was discovered between reading proficiency and the visual factors appraised in the study. Nevertheless, for the sake of ocular comfort and visual efficiency, every child should have corrections where indicated. (3) Frequency of need for lens correction was found to be relatively high (62.4 per cent). Because this percentage is surprisingly high, the findings should be checked.

Certain questions may be raised concerning the study: (1) Great emphasis is placed upon the dynamics of seeing, or seeing as learned. Yet the authors indicate that about two-thirds of their group of fifth-graders should wear glasses. Perhaps the glasses are to serve only as crutches until training becomes effective although no re-education program for these children is specified. If the functional emphasis upon seeing as a psycho-physiological process can be defined in terms that lead to practical application, a complete statement of the re-education program would be helpful to the reader. (2) "In almost every instance investigators agree that there seems to be a relationship between visual findings and readiness for sustained attention in reading activities, though some studies fail to demonstrate it." Probably

XUM

the agreement is not as universal as implied here. It is now considered by many that physiological visual deficiencies as determinants of reading proficiency have been overemphasized. Nevertheless, the maintaining of hygienic vision need not be justified in terms of whether or not visual status is related to reading proficiency. (3) Emphasis is placed upon use of screening tests for identification of children who should be referred to eye specialists. Validation of the criteria employed to identify referrals is not given. In fact, it is stated that "validation of referrals is a problem for the refractionist and not those using the screening tests." Also, since the ophthalmological records are available, one would like to know if the screening tests missed any children in need of lens correction. (4) Many readers will feel that analyses and discussions are frequently incomplete. Although the authors had planned not to give interpretations, readers will be disappointed by the omission.

The comprehensive data reported in thi atise will promote a better understanding of visual problems of s ol children. Similar studies at other grade levels are needed.

MILES A. TINKER.
University of Minnesota.

R. C. OLDFIELD. *The Psychology of the Inter w.* London: Methuen and Company, 1941, pp. 144.

This is a convenient little essay of one hundred forty pages on the principles and practices of interviewing. The interview is defined as conversation between two or more individuals for a definite purpose among which the author mentions the collecting of information, making arrangements, selection and rejection of applicants for employment, the assessment of intellectual and temperamental qualities, diagnosis and treatment in medicine, and the adjustment of grievances and difficulties. Considerable emphasis is placed on the fact that the interviewer bases his judgment on the attitudes displayed by the candidate before him. The setting and conduct of the interview are given interesting treatment and there is a chapter on the psychological process in forming judgments. A very valuable chapter concerns the board interview with a small section devoted to multiple interviews. The discussion of the group interview seems to be omitted.

On the wrapper which is folded around the book, it is stated, "It can be claimed that this is the first scientific study of the interview." This may be the publisher's statement and one cannot be certain whether or not this claim emanates from the author. We also learn

XUM

from the foreword by Dr. C. S. Myers that this book is a product of a research studentship under the National Institute of Industrial Psychology which the author held for fifteen months. This book is evidently a product of this studentship.

One can dispute the claim on the wrapper that this is a "scientific study of the interview," in the first place, and that it is "the first" study, in the second place. I find that except for a few comparisons in the chapter on the board interview there is no reference to data which might have been collected during the period of the study which would serve as the basis for the conclusions reached. As a matter of fact, a considerable part of the discussion would appear to consist of general observations which frequently have been made concerning the art and process of interviewing. That this is the first study of the art of the interview is certainly contrary to fact. Oldfield does not include a bibliography or any reference to the work of other students of the interview, but it is difficult to believe that a person could have spent fifteen months in studying the psychology of the interview without having discovered during that time the work by Bingham, Young, Strang, Symonds, and many others.[1]

Although the author states that an interview may be used for a number of purposes, it is not clear throughout the essay exactly what kind of interview he is talking about. One would assume during most of the discussion that he had in mind the interview for the "selection and rejection of applicants for employment." It is well known that the process of interviewing varies according to the use that is being made of it. This functional aspect of interviewing seems to have been overlooked by the writer. I find that this author thinks of the interview largely in terms of question and answer. For instance he says, (page 72) "A third advantage of routine is that by it the interviewer's mind is freed, to some extent, from the task of formulating his questions and considering his next move." The implication here is that the interviewer's task is largely that of formulating questions, noting replies, and observing the responses of the candidate. This is a very limited concept of the interviewing process.

[1] Bingham, W. V., and Moore, B.: *How to Interview.* Harper & Brothers, 1931, 1941.

Young, P. V.: *Interviewing in Social Work.* McGraw-Hill Book Co., 1935.

Strang, R. M.: *Counseling Technics in College and Secondary Schools.* Harper & Brothers, 1937.

Symonds, P. M.: *Diagnosing Personality and Conduct.* D. Appleton-Century Co., 1931.

There are certain statements in the book which would seem to be much more a matter of personal opinion than the result of a scientific study. For instance, it is stated, (page 54) "Some interviewers find that movements and posture of the whole of the candidate's body and not merely of the upper half are significant for judgment. For this reason, there may be some advantages in making the candidate sit so his legs are visible. In the second place in order to observe the face of a female candidate, removal of the hat may be necessary."

Notwithstanding the above critical comments, this book takes its place among others of its kind as a compact and well-phrased statement of the art and process of the interview. The convenient size and length of the book commends itself to the use of students who wish to gain a quick survey of the interviewing process.

PERCIVAL M. SYMONDS.
Teachers College, Columbia University.

MANDEL SHERMAN. *Basic Problems of Behavior.* New York: Longmans, Green and Company, 1941, pp. 417.

Basic Problems of Behavior will be easily enough recognized by students who are familiar with Dr. Sherman's *Mental Hygiene and Education* as a revision and elaboration of that earlier volume. In its present form, however, they will find it more useful for a first course in mental hygiene than was the original volume. Motivation and frustration is considered early in the volume and the experimental and practical phases of this problem are fairly adequately discussed. Other topics discussed in the ten chapters include theories of personality, common mechanisms of adjustment, the measurement of personality, attitudes, conflicts, delinquent behavior, the neuroses, and mental abnormalities. The general style of writing in this book is intelligible, concise, clear, direct, and adequate for the purpose. The point of view is fairly eclectic and the general emphasis fairly practical.

However, it is not the kind of a book that one dare put in the hands of a phobiac; at least, judging by one experience that the reviewer had, all the phobiac will get out of it is that Dr. Sherman says on page 379 that fears and phobias occur in less intelligent persons, and that somewhere in the beginnings of the book he says that phobiacs are next to impossible to cure. What Dr. Sherman really says is that fears and phobias are more likely to occur in less intelligent persons, and what he says about the cure of phobiacs is that case histories of treated patients show that the frequency of cure of phobiacs is quite low. For

the realistic consideration of such problems caution is quite warranted but for influencing a phobiac or for inspiring a fearful person to want to learn how to overcome his fears or phobias, it certainly is not the best selection in the early stages of treatment. However, since the book was written, not for neurotic individuals, but for class purposes and for more intelligent readers familiar with psychology, the fact that it will not serve a neurotic well in early stages of treatment is hardly a relevant criticism. The reaction of such a patient, however, shows the need for a book that is accurate and also more realistically encouraging. The chief use for this book that the reviewer can see is for a first course in mental hygiene, and for that purpose it has advantages that other books on the market do not have. It is not over-weighted with neurological and physiological considerations or with sensory psychology and related problems that are not generally significant or relevant in a consideration of problems as they appear in a psychological clinic or in the office of a psychologist or a psychiatrist.

Students of differential psychology would probably disagree with Sherman's statement on page 114 to the effect that "systematic interest in personality developed with the studies of Wundt and Galton, who were interested in differential psychology, such as studies of individual differences." Most psychologists think of Wundt as being almost opposed to the study of differential psychology and would not tie up Wundt's work with Galton's early work, as far as individual differences are concerned. In fact, Columbia University has it that Cattell broke away from Wundt to study with Galton in order to study individual differences. This, however, is nothing more than a minor error in the history of psychology or in an interpretation of an historical fact. Another minor error that has crept into the volume, which merely proves it is a man-made world and all contributors are males unless they describe themselves as women, is in reference to C. H. Towne as "he" on page 61. Also, in a consideration of the emotions, the author perhaps is more academic than one would expect him to be and gives insufficient attention to actual emotional adjustments.

In spite of these minor errors, the book taken as a whole represents a splendid integration of relevant experimental knowledge and first hand familiarity with therapeutic techniques that practicing psychiatrists as well as educational psychologists will find deserving of their reading time.

H. Meltzer.

Psychological Service Center, St. Louis, Missouri.

XUM

XUM

A New and Up-to-date

COURSE OF STUDY IN ENGLISH

Covering the teaching of oral and written English, usage and technicalities, and reading (both the recreational and study types) for

PRIMARY GRADES

INTERMEDIATE GRADES

JUNIOR HIGH SCHOOL

based, as far as possible, upon the best that experimentation and expert opinion have produced. Prepared by the Supervisors and Teachers of

BALTIMORE COUNTY

under the direction of Dr. R. L. Lyman.

Subjects Covered

Reading: (Grades I to IX) Recreational—reading for pleasure, with emphasis upon excellent literature in both prose and poetry; for study—(1) silent reading to gain information and to improve one's skill in comprehension; (2) oral reading to give information to others and to improve one's ability in interpreting what is read.

Expression, oral and written: (Grades I to III) The expression in daily use by the pupils for such purposes as telling stories, talking together, answering questions, relating experiences.

Communication, oral and written: (Grades IV to IX) The composition in daily use by the pupils for such purposes as conversation, telling stories, answering questions, writing letters, giving directions, reporting experiences, participating in group discussions.

Language: (Grades I to III) Items dealing with the mechanics of oral and written expression, capitalization, and punctuation.

Spelling and Handwriting: (Grades I to III) The items dealing with techniques of study in spelling and handwriting.

Technicalities: (Grades IV to IX) Items dealing with the mechanics of communication, namely, sentence work and grammar, usage, capitalization, punctuation, and spelling.

Part One, Grades I to III. Part Two, Grades IV to VI. Part Three, Grades VII to IX.

Price: $1.42 per section postpaid. $3.83 per set postpaid. Special discounts on quantity orders.

WARWICK AND YORK

Publishers BALTIMORE, MD.

XUM

Educational Psychology

BY

Four Widely Experienced Men in the Educational Field

ARTHUR I. GATES

ARTHUR T. JERSILD T. R. McCONNELL ROBERT C. CHALLMAN

This work is especially planned to be of maximum usefulness to the teaching profession. It presents, with authoritativeness and completeness, the wealth of assistance the psychologist is now able to give the educator—understanding of child development and of learning processes; methods of guiding and stimulating learning and development; analysis and measurement of intelligence, aptitudes, and abilities or disabilities; standards and safeguards of mental health. The book is one which will help the teacher see his professional activities with deeper insight and in broader relationships, and carry forward his work with more competence and satisfaction.

To be published in February.

$3.00 (*probable*).

The Macmillan Company • 60 Fifth Avenue New York

THE MAPLE PRESS COMPANY, YORK, PA.

XUM

ORK, PA.

XUM